

THIS WEEK

EDITORIALS

PARTICLE SPIN What to do if we fail to find the Higgs boson **p.6**

WORLD VIEW The cause and effect of contraceptives and HIV **p.7**

COMPUTING Pioneer of AI John McCarthy dies **p.11**



The mind's tangled web

Efforts to elucidate how genes and the environment shape the development of autism, although making progress, still fall far short of their goal.

Among psychiatric disorders, autism has received particularly strong support from government and philanthropic funders in recent years. And that investment has paid scientific dividends, above all the uncovering of genetic clues to underlying mechanisms for the disorder. But, as discussed in this special issue and in a web collection of content published this week in other Nature journals (see nature.com/autism), those developments, although pointing a way forward, have themselves revealed just how small a distance we have travelled towards a fuller mechanistic understanding.

Context, as always, is important. First, the object of study: it is clear that the diagnostic criteria for autism need to be refined and expanded, and that there is a spectrum of autistic disorders. Second, although most parents struggling with children with autism would jump at the chance to mitigate or cure the symptoms, it is not appropriate to think of autism solely as a disorder needing treatment. As psychiatrist Laurent Mottron vividly describes on page 33, people with autism bring particular talents to many professional settings, including the scientific laboratory. And that perspective plays into a range of ethical implications for the pursuit of biomarkers for the condition (P. Walsh *et al. Nature Rev. Neurosci.* **12**, 603–612; 2011). Third, the growth in the prevalence of autism can be explained only partly by changes in diagnostic practice — about 50% seems to be genuine, or at least unexplained (see page 22).

Everyone agrees that autism stems from a disruption of brain development caused by a combination of genes and environment. Since the 1970s, it has been known from studies of twins that there is a high, but not complete, degree of heritability. In recent years, well-funded and coordinated efforts, coupled with advances in technology, have led to large-scale studies of unprecedented statistical power, producing impressive data on the genetics of the condition. But those data have confirmed only that the answer is elusively complicated. With the exception of a few rare disorders, such as fragile X or Rett syndrome, which lead to forms of autism, no disruption of an individual gene, or set of genes, can reliably predict the condition. An emerging story is that the culprits could include any one of many extraordinarily rare genetic variations, and that a systems approach will be important in understanding regulatory hubs, for example (M. W. State and P. Levitt *Nature Neurosci.* <http://dx.doi.org/10.1038/nn.2924>; 2011).

Nonetheless, progress is being made in tracking the neurobiological effects of these genetic variations. Recently, mice carrying mutations in candidate genes have been produced, and have been found to show behaviours reminiscent of autism — such as a lack of interest in socializing with other mice, repetitive grooming and anxiety. These mice all have alterations in brain structure and function, and will undoubtedly be useful for testing hypotheses about the relationship between various brain circuits and autism-associated behaviour (see, for example, J. L. Neul *Nature Med.* **17**, 1353–1355; 2011). Along with these mice, cellular models derived from patients carrying certain mutations also hold promise for testing molecular hypotheses and therapies.

Encouraging as these efforts are, they cover just half of the gene–environment equation and thus will at best only ever yield part of the solution. It is widely agreed that environmental factors, through direct neurobiological mechanisms or interactions with genes, could interfere with neural development to cause autism. Many factors have been proposed, including maternal infection during pregnancy. But none of these candidates has yet been convincingly established, nor have their biological links to autism been tested mechanistically.

“The growth in the prevalence of autism can be explained only partly by changes in diagnostic practice.”

Clearly, nailing down a given influencing factor from our incredibly complex environment is no trivial task, possibly harder than identifying variation in the genome. It will take very large and expensive long-term studies just to narrow down the possibilities.

Some, such as the US National Institutes of Health's EARLI study (see page 22), are already under way and could yield not only valuable material for studying gene–environment interactions, but also potential mechanistic leads for biologists to pursue.

Although many scientists and funding agencies have focused their resources on apparently tractable questions in genetics and neurobiology, the attention of the public and mainstream media has been repeatedly drawn to environmental hypotheses regardless of their scientific strength, such as the discredited link to vaccinations (see page 28). It is essential that agencies and philanthropists are not similarly distracted by public mood, just as it is crucial that scientists working on possible environmental influences take great care to communicate the results of their studies properly. The need to elucidate the true environmental influences on autism is a priority, and, as our collection of articles displays, social sciences, psychology and neurobiology all have their roles. ■

More in Montreal

Momentum builds for ozone treaty to take on greenhouse gases.

For more than two decades, the 1987 Montreal Protocol has served as a shining example of how to get things done on the environment in the international arena. By banding countries together to preserve Earth's shield against harmful ultraviolet rays, the agreement has already eliminated many ozone-depleting substances and should see off most of the rest by 2030. And in doing so, it has done more to reduce greenhouse-gas emissions than the 1997 Kyoto Protocol, which was signed expressly for that purpose. It is equally

clear that Montreal could do more on the climate front — if a handful of countries would simply put politics and turf battles aside.

The next opportunity for progress will come when ozone negotiators gather later this month for their annual meeting in Bali, Indonesia. On the agenda again are proposals to regulate a potent and increasingly important class of greenhouse gases known as hydrofluorocarbons (HFCs), and once again China, India and, to a lesser extent, Brazil are expected to block any real action. It would seem that the best outcome would be a strong vote on yet another non-binding resolution to consider the issue again next year. In this sense, little seems to have changed since the summer of 2009, when *Nature* first recommended that Montreal delegates accept this new challenge (see go.nature.com/yraznh). But there is still reason for hope.

The official line on why obstructing countries at the Bali meeting will oppose HFC regulation under Montreal is bureaucratically flawless: HFCs do not destroy ozone and so are the responsibility of parallel climate negotiations, which will reopen in Durban, South Africa, a week later. Opposing countries also fear a precedent-setting shift of climate regulations out of the Kyoto Protocol — a treaty that developing countries support because it puts the onus on industrialized nations — and into a framework in which action is mandatory for everybody. The Montreal Protocol itself has become a bargaining chip in the climate negotiations.

This is unfortunate, as the logic for tackling HFCs under the Montreal Protocol is powerful indeed. After all, the gases were introduced to replace the ozone-eating compounds outlawed by the treaty. And they could be regulated using the same tools that have helped the agreement to promote an orderly transition to less damaging chemicals. Moreover, because HFCs can be thousands of times more effective at trapping heat than carbon dioxide, they unbalance the carbon markets on which Kyoto relies. (The most notorious example, HFC-23,

is a waste product of the production of Teflon and other materials. Some companies in India and China make more money from destroying HFC-23 under the Kyoto Protocol's Clean Development Mechanism than they do from selling chemicals.)

“The Montreal Protocol itself has become a bargaining chip in the climate negotiations.”

In effect, industrialized countries are paying way over the odds to destroy HFCs. Recognizing this, the European Commission has already moved to ban any new HFC-23-destruction projects from its carbon trading scheme.

By contrast, under Montreal Protocol rules, industrialized countries would lead the development of alternative chemicals and phase out the most potent greenhouse gases, while putting money on the table to help developing countries follow suit. It is a fair system based on actual costs, and it works.

The Federated States of Micronesia first proposed targeting HFCs under the Montreal Protocol as part of a simple climate agenda that could be advanced quickly while the world deliberates over the harder problem of carbon dioxide. The United States, Mexico and Canada joined the cause, eventually followed by the European Union. By 2009, some 41 countries had offered their support for the idea. That number increased to 91 countries in 2010, and it could go even higher this year.

Montreal will celebrate its twenty-fifth anniversary in 2012, and a fitting way to honour its accomplishments would be to expand its mission. Momentum continues to build, leaving China, India and Brazil increasingly isolated on this issue. Each of these nations has been critical of industrialized countries for not doing enough to curb greenhouse-gas emissions, and rightly so. But in this case, they have only themselves to blame. ■

Collision course

What will scientists do if they fail to find the Higgs boson?

Largely unnoticed at the weekend, the Large Hadron Collider (LHC), the world's most powerful particle accelerator, ended one hell of a run. Since March, protons had been whizzing more-or-less continuously around the collider's 27-kilometre ring, located at CERN, Europe's premier high-energy laboratory on the French–Swiss border near Geneva. At four points around the ring, protons collide at near-light-speed. The energies released are high enough to briefly create heavy particles, and physicists have been hoping for one above all others — the Higgs.

The Higgs boson is a missing bolt in the standard model of particle physics. In popular parlance, it is known as the particle that gives all other particles mass. However, physicists like the Higgs (and its associated field) because the particle provides an elegant way to combine electromagnetism with the weak nuclear force. Such a step would create a single electroweak theory that is stronger than its two parts. It would also consolidate the four known forces into three — bringing researchers closer to an ultimate theory of everything. There are good reasons to hope that the Higgs is there, but nothing in nature requires it. Already in August, physicists saw one possible signal sink from view.

With the end of this year's run, the two largest detectors at the LHC have each collected six inverse femtobarns (about 42×10^{13} collisions worth) of data. Crucially, that is enough to tell whether the Higgs takes the form that physicists believe it must for it to give mass and unify forces.

In other words, if the Higgs does not exist, then the world will probably be told so within weeks. (Or at least that scientists are 95% sure that it does not exist.)

A Higgs no-show would certainly create a public-relations headache for CERN, for the field of high-energy physics and possibly for science itself. For better or worse, the boson is popularly referred to as the ‘God particle’. Scientists forced to admit that they have seen no sign of it would no doubt face the wrath of non-believers — the kind that asks where decades of hard graft and taxpayers' billions have gone.

In September, CERN's council discussed a report on ‘The scientific significance of the possible exclusion of the SM Higgs boson in the mass range 114–600 GeV and how it should be best communicated’. The public version of the document, available online, emphasizes that the failure to find the Higgs would be just as exciting as a discovery. Privately, discussions are under way on whether the lab should announce that a negative result ‘excludes’ the Higgs, which sounds final, or merely ‘disfavours’ it — as 95% leaves a bit of wriggle room.

Spin aside, scientists are unlikely to panic. If the predicted Higgs is not there, they will stick to their plans and collect additional data in next year's run. If that does not turn up a signal, then they will have to wait for more than a year for the LHC to undergo an upgrade that will allow it to run at twice its current energy. At these higher energies, the standard model becomes much less predictable, and researchers will either find something new, or they will greatly improve their understanding of the existing theory. The collisions that happen in the LHC are complex, as is the analysis, and this extra effort could take years.

If high-energy physicists eventually find something new, then the field may flourish. If they only refine the existing models, it may eventually fade. Regardless, the LHC does what experiment must: it puts a good hypothesis to the test. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunq



Scientists and the media must give a balanced view

A reported link between hormonal contraception and HIV infection deserved a more critical look, says James Shelton.

Women using hormonal contraception (HC) may face an increased risk of contracting HIV, according to a paper published online last month by *The Lancet Infectious Diseases* — especially women using a popular injectable contraceptive. *The New York Times* proclaimed that the injectable “appears to double the risk the women will become infected with HIV”. Other alarming media accounts followed. But what is the evidence? And how well was it reflected in the resulting media coverage?

The study suggested not only an increased risk for women using HC, but also increased transmission to their partners (R. Heffron *et al. Lancet Infect. Dis.* [http://dx.doi.org/10.1016/S1473-3099\(11\)70247-X](http://dx.doi.org/10.1016/S1473-3099(11)70247-X); 2011). And it implicated both injectables and oral contraceptives, although the latter did not reach statistical significance. Whether HC influences HIV risk is a serious concern, and has been the subject of numerous studies. But these studies have been observational and not randomized, and thus potentially biased by who chooses to use HC.

A classic way to interpret such evidence is to use causality criteria laid down by British epidemiologist Austin Bradford Hill. Applying some of these (in bold) I find the evidence far from persuasive.

Consistency. A notable number of studies have found no increased risk; others found only sporadic increases in disparate subgroups.

Strength of association. The reported twofold risk could be considered moderately supportive. However, because the number of infections among HC users and their partners were few, the wide confidence intervals included an increased risk of only a few per cent. And one alternative analytical approach was not significant. Furthermore, the study was a secondary analysis of data from an HIV-prevention trial addressing a different question, making it susceptible to additional bias because chance occurrences are more likely to be found and published.

Absence of alternative explanations. One plausible alternative is that couples who don't use HC are more likely to use condoms. The authors did attempt to control for this, but condoms were actively promoted and condom use seems greatly over-reported. The reported level of unprotected sex (about 10%) is inconsistent with HIV and pregnancy rates.

Biological plausibility. There seems to be no clear mechanism to explain how both forms of HC could increase the risk of both acquisition and transmission. The study found a small rise in genital viral shedding, the team's putative and seemingly necessary mechanism for transmission, but this was too small to explain the increased risk. Paradoxically, finding the same twofold risk for acquisition and transmission for both injectables and oral contraceptives actually strengthens the likelihood that reduced HIV exposure in the non-HC group can explain all the findings.

This assessment is open to other interpretations, but why so little critical thinking in media reports? Let's look at the role of various actors.

Study authors: it taxes us to criticize our own work. However, authors should lay out critical issues for examination. The authors of the research paper did include some problems with the methodology and mentioned previous inconsistent findings. However, a prominent panel included in their paper, labelled “Research in Context”, cited only studies that found increased risk of acquisition. A more balanced representation was clearly in order. The issues of condom use and the increase in genital viral shedding that was seemingly inadequate to account for the transmission risk were not discussed, even though six authors co-authored another paper, from essentially the same study population, quantifying shedding's overall positive relationship with transmission.

Comment authors: it is crucial that accompanying comment articles in scientific journals lay out issues in a balanced, even critical, manner. Indeed, the title of the comment published alongside this study — “Hormonal contraception and HIV: an unanswered question” — made clear the uncertainty. And it did a good job on some issues, including the secondary nature of the analysis. But it overly emphasized those studies that reached a similar conclusion to the original study, and was uncritical on biological mechanisms, including viral shedding.

Journal: the journal's press release asserted the study's positive findings, and the quandary for contraceptive choice — omitting entirely any limitations or conflicting findings. In addition, journals frequently choose reviewers — who

probably approve publication — to write accompanying comments. Is that the best way to provide critique and perspective?

The New York Times: this led with the finding of increased risk and dwelled extensively on the implications of the conflict between prevention of unintended pregnancy and prevention of HIV. Readers had to reach paragraph nine for a description of the study. The only specific limitation of the study described was left to paragraph 16, and the sole reference to other human studies (paragraph 13) stated only that “at least two other rigorous studies” had found increased acquisition risk — no mention of work showing the opposite. Condom use and biological plausibility were mentioned only in passing. The article essentially presented the increased risk as established, and focused on the conflict it presented.

Such a conflict piques interest, but ill-serves what remains a complicated and serious issue. We deserve better, from all involved. ■

James D. Shelton is science adviser, Bureau for Global Health, US Agency for International Development, Washington DC, USA. The views expressed are not necessarily those of USAID.
e-mail: jshelton@usaid.gov

APPLYING HILL'S
CRITERIA, I FIND
THE EVIDENCE
THAT HORMONAL
CONTRACEPTION
INCREASES HIV RISK
FAR FROM
PERSUASIVE.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/fifyay

SEVEN DAYS

The news in brief

RESEARCH

Higgs hunt halts

The Large Hadron Collider (LHC) has finished smashing protons for the year. The particle accelerator at CERN, Europe's premier high-energy physics laboratory near Geneva, Switzerland, ended its 2011 run on 30 October, after about 180 days and 400 trillion collisions. Physicists are still trawling through the data for evidence of the Higgs boson — the particle thought to give other fundamental particles their mass. Meanwhile, the LHC is moving on to collisions between lead ions, and between lead ions and protons. Proton–proton collisions will start again in March 2012. See also Editorial, page 6.

Archives made free

Benjamin Franklin's account of his electric kite experiment (in 1752) and Isaac Newton's first scientific paper (1671) are among almost 70,000 historical scientific papers now freely accessible online, after Britain's Royal Society opened up its journal archive. The archive goes back to 1665, when *Philosophical Transactions* first appeared. From 26 October, all papers published more than 70 years ago are free to view, and the archive is fully searchable. See go.nature.com/wka6bw for more.

Climate craft

In a small but notable success, on 28 October NASA launched a polar-orbiting satellite to collect data for climate researchers and weather forecasters. The craft marks a new beginning for the United States' troubled programme to create the next generation of polar-orbiting satellites (see *Nature* 477, 509; 2011). It may also help to stave off a gap in data collection expected when existing polar-orbiting



AYSÉN REGIONAL GOVERNMENT / AP

Hudson volcano threatens to erupt

Chile's ice-covered Hudson volcano awoke from a 20-year slumber on 26 October, raising fears that a major eruption could be due. As *Nature* went to press, three vents were spewing out steam and ash from the volcano (pictured) in the Aisén Region of northern Patagonia, and

geologists were monitoring earthquake swarms. Mount Hudson's last eruption, in 1991, was one of the largest of the twentieth century, although it was overshadowed by an even bigger eruption at Mount Pinatubo in the Philippines, just a few months earlier.

satellites shut down in 2015. However, faced with funding shortfalls, NASA is already falling behind on a follow-up satellite, scheduled for launch in 2016. See go.nature.com/mkh4wa for more.

lowers the cost per genome to no more than \$1,000 and requires an accuracy of no more than one error per million bases. See go.nature.com/5jx395 for more.

New York campus

A contest to develop a science and engineering campus in New York City attracted seven bids by its 28 October entry deadline, according to city mayor Michael Bloomberg. Winners of the competition will be granted free land in the city and US\$100 million in seed money. Bloomberg — who said he might choose multiple winners — hopes that the project will help the city to become a technology-science hub, akin to Silicon Valley in California. Fifteen universities are among the bidding consortia, including Stanford

University in California; Cornell University in Ithaca, New York; the University of Toronto in Canada; the Technion–Israel Institute of Technology in Haifa; and the Indian Institute of Technology Bombay.

Saving San Raffaele

Despite accumulating debts of €1.5 billion (US\$2.1 billion), the San Raffaele Scientific Institute in Milan, Italy, will not have to file for bankruptcy. A restructuring plan put together by the board of the prestigious biomedical research and hospital complex was accepted by a Milan court on 27 October. The plan is backed by a rescue package led by the Vatican Bank (see *Nature* 478, 296–297; 2011). However, some recently appointed board members are connected with the rescue

FUNDING

Sequencing prize

The non-profit X Prize Foundation is offering US\$10 million to any team that can accurately sequence 100 genomes from 100 centenarians in 30 days, starting on 3 January 2013. The competition is a refocusing of the Archon Genomics X Prize, which in 2006 laid down the challenge to sequence 100 genomes in 10 days at less than \$10,000 per genome, but netted no winners. The new competition

AP consortium. Worried by potential conflicts of interest, the Milan court said that it would appoint a committee to oversee the restructuring, and told the board to remain open to new rescue offers.

Jackson lab grows

The Jackson Laboratory, a medical research centre based in Bar Harbor, Maine, will set up a major satellite facility for personalized medicine and systems-genomics research near the University of Connecticut Health Center in Farmington. The move was confirmed on 26 October, when Connecticut's state legislature approved US\$291 million in bonds to fund the new laboratory. The lab had tried for more than a year to site its facility in Florida, but the state would not provide funding.

US cuts off UNESCO

The United States says that it will stop all payments to the United Nations Educational, Scientific and Cultural Organization (UNESCO) after the Paris-based organization voted on 31 October to give full membership to the Palestinians, who are seeking statehood. Existing US legislation says that no UN body making such a move can be funded. In total, the United States was due to contribute \$80 million to UNESCO this

financial year, around 22% of the agency's total direct funding from members.

BUSINESS

Solar subsidy cuts

The United Kingdom is cutting incentives to produce solar power. On 31 October, the government said that feed-in tariffs for solar energy (the price that an electricity utility company must pay to generators) should be reduced by more than 50% because too many solar installations were being planned. The new tariffs are subject to consultation, but would bring down government support roughly in line with that in Germany, which announced its own 15% tariff cuts on 27 October. France, Spain and Italy have all cut their solar tariffs in the past year.

PEOPLE

AI pioneer dies

John McCarthy (pictured), a visionary computer scientist who coined the term 'artificial intelligence', died on 24 October, aged 84. In 1958 — two years after organizing the first conference on artificial intelligence — McCarthy invented the computer programming language LISP. He spent most of his career at Stanford University in California, and developed ideas for



computer time-sharing. "He was always focused on the future. Always inventing, inventing, inventing," said Stanford computer scientist Ed Feigenbaum.

Weapons lab head

Penrose 'Parney' Albright will be the eleventh director of the Lawrence Livermore National Laboratory in Livermore, California, the lab announced on 27 October. Albright, a doctoral physicist who has spent much of his career working in defence and homeland security, replaces George Miller, who spent six years as director, from 1 December.

US solar woes

One of the United States' largest solar-panel companies, First Solar, abruptly and unexpectedly lost its chief executive on 25 October. Robert Gillette has left after two years at the helm of the firm, based in Tempe, Arizona. First Solar is a leader in the thin-film photovoltaic market, but its share price has dropped

COMING UP

4 NOVEMBER

Six men who have spent 520 days cooped up at the Institute of Biomedical Problems in Moscow, simulating the isolation of an expedition to Mars, 'return' to Earth. go.nature.com/uupu8s

8 NOVEMBER

Russia's Phobos-Grunt mission — which it is hoped will return a sample of soil from the Martian moon Phobos — is scheduled to launch, together with China's first probe to Mars, Yinghuo-1.

8-9 NOVEMBER

The European Gender Summit in Brussels discusses how to support women's scientific careers, and how to tackle gender factors in research. www.genderinscience.org

by 65% in 2011. This year has seen several bankruptcies of solar firms after the price of solar panels plummeted in response to weak demand and a surge in low-priced modules from China.

Psychology fraud

A preliminary investigation has revealed massive fraud by a prominent social-psychology researcher, Diederik Stapel, who faked data in at least 30 published studies. The report was released on 31 October by a committee from Tilburg University in the Netherlands, together with committees at the universities of Groningen and Amsterdam. In September, Tilburg suspended Stapel on suspicion of research misconduct. See page 15 for more.

NATURE.COM

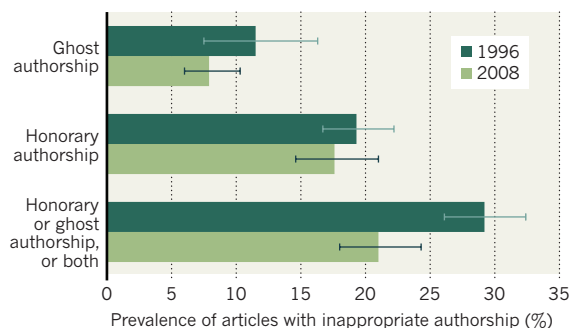
For daily news updates see: www.nature.com/news

TREND WATCH

Inappropriate attribution of authorship plagues research papers, according to a study published on 25 October (J. S. Wislar *et al.* *Br. Med. J.* 343, d6128; 2011). A survey of 6 high-impact biomedical journals found that 21% of articles published in 2008 had either honorary authors (who were named as authors but hadn't contributed enough to take responsibility for the work) or ghost authors (who had contributed to the work but weren't named).

REAL AUTHORS WANTED

Ghost and honorary authorship in high-impact biomedical journals has declined only slightly since a 1996 survey*.



*Survey of authors of 896 articles in *Annals of Internal Medicine*, *Journal of the American Medical Association*, *Lancet*, *Nature Medicine*, *New England Journal of Medicine*, *PLoS Medicine*

NEWS IN FOCUS

ETHICS Investigation finds that psychologist committed massive fraud **p.15**

EARTH SCIENCE Probing the fault that sparked the Japanese tsunami **p.16**

ASTRONOMY Giant telescopes face giant funding hurdles in the United States **p.18**

SPECIAL ISSUE What science knows about autism **p.21**



and put doctors who offer *in vitro* fertilization (IVF) in a dubious legal position, because not all embryos created during fertility treatment survive the procedure.

"This is a dangerous and extreme government intrusion into women's health, women's rights and families' health," says Stan Flint, a consultant to Mississippians for Healthy Families, based in Jackson, which opposes the amendment.

Similar propositions have been put to voters in the United States twice before — during statewide campaigns in Colorado, where the personhood movement first emerged as a strategic challenge to abortion laws. But in both 2008 and 2010, personhood initiatives were roundly defeated, respectively winning only 27% and 29% of votes cast.

Mississippi could be very different. With candidates from both parties, the state's attorney general and a phalanx of religious groups all endorsing Initiative 26, there is a widespread sense that the measure could pass. Such a victory would undoubtedly boost efforts to pass personhood initiatives in several other states, from Oregon to Florida (see 'States of play'), where similar measures are set to appear on ballots in November 2012. Ultimately, Personhood USA hopes to win initiatives in at least two-thirds of the 50 states. Those states, in turn, could then require Congress to put forward a personhood amendment to the US Constitution.

Even if that strategy falls short, victory in just a few states would shift the balance in the ongoing US debate over the handling of human embryos, inviting future restrictions or conditions on federal funding for stem-cell research. "I think the field should be worried," says Alta Charo, an expert on law and bioethics at the University of Wisconsin-Madison. "They are still being held hostage by the abortion debates."

The Mississippi vote itself will have little direct impact on human embryonic stem-cell research, because the state is not a major player in the field. The potential threat to reproductive technology is more immediate.

The personhood movement says claims that the initiative would restrict fertility treatments and birth control are fear mongering. "It's not going to ban IVF and contraception. It's only going to ban procedures and pills that kill a person," says Mason. "If it kills a person it should be illegal." ▶

Supporters of Initiative 26 hope to outlaw abortion by changing the definition of person to include embryos.

POLITICS

Mississippi to vote on 'personhood'

Ballot measure would give fertilized eggs human rights.

BY SUSAN YOUNG

"When do you believe life begins?" Johnny DuPree, Democratic candidate for governor of Mississippi, asked during a public debate on 14 October. The question was rhetorical, and DuPree's answer — not a surprise in one of the most socially conservative US states — was the same as that of his Republican opponent: "I believe life begins at conception."

On 8 November, Mississippi voters will not only decide who should lead the state, but also indicate whether they agree with the candidates about the status of embryos. The Initiative 26 ballot measure proposes to amend the state's constitution to redefine 'person' as

"every human being from the moment of fertilization, cloning, or the equivalent thereof". If approved, the amendment would effectively bestow human rights on fertilized human eggs, making abortion illegal in the state in most, if not all, circumstances.

"The unborn child in the womb is scientifically proven to be a human being, and when it comes down to it we are a human-rights organization," says Jennifer Mason, communications director for Personhood USA, a national advocacy group based in Arvada, Colorado, and one of the initiative's backers.

By defining personhood so broadly, the measure would also have an impact beyond abortion — for example, it could rule out research using human embryonic stem cells

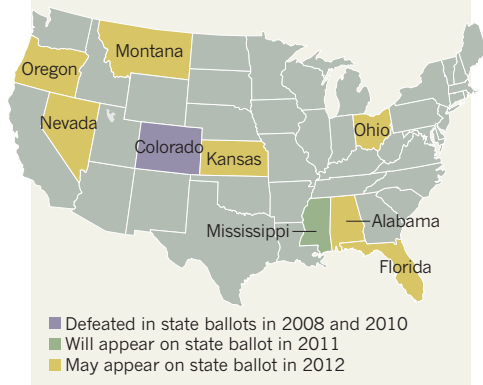
► But opponents say that the wording of the amendment means that it could interfere with established medical practices. “If personhood begins at fertilization, then we have to talk about IVF and birth control,” says Jonathan Will, director of the Bioethics and Health Law Center at Mississippi College in Jackson. Some forms of contraception, such as the intrauterine device and emergency hormonal contraceptive pills, prevent fertilized embryos from implanting in the uterine wall and so could be considered illegal under the amendment, experts say.

Fertility doctors add that the measure could hamper IVF and endanger the would-be mother and her offspring.

To give patients the best chance of pregnancy, doctors typically fertilize 8–10 eggs and implant only the one or two embryos that seem most vigorous. The rest are stored or discarded. If a doctor is forced to implant all fertilized eggs to avoid prosecution, then the patient is more likely to have multiple pregnancies, which can be risky for her and the fetuses. Yet limiting the

STATES OF PLAY

‘Personhood’ initiatives or their equivalents, which aim to grant human status at the moment of fertilization, are becoming an increasingly familiar feature of state-wide elections in the United States.



number of embryos created for IVF to only the number of children desired reduces the chance of success and increases the likelihood that women will have to undergo the difficult and

expensive procedure more than once.

Proponents of Initiative 26 point out that other countries have already legally limited the number of fertilizations in IVF. In Italy, for example, a law introduced in 2004 limits doctors to fertilizing only three eggs and requires all resulting embryos to be implanted. Yet studies suggest that the law has reduced the success rate of IVF and increased the number of triplet pregnancies (P. E. Levi Setti *et al. Fertil. Steril.* 90, 1081–1086; 2008).

As the campaign for Initiative 26 heads into its final days, voters have been bombarded with commentaries, blogs, YouTube videos and public rallies on both sides of the debate. Flint acknowledges that, at this point, the defeat of the Mississippi initiative would be a turnaround, but an increasingly vocal opposition movement has thrown predictions of an easy victory for the initiative into question. “Starting from a dead stop at two months out, we have put together a major campaign,” says Flint. “The momentum has swung strongly towards the opposition to this amendment.” ■

PSYCHIATRY

Mental-health guide accused of overreach

Dispute grows over revisions to diagnostic handbook.

BY HEIDI LEDFORD

Psychologist David Elkins had modest ambitions for his petition. He and his colleagues were worried that proposed changes to an influential handbook of mental disorders could classify normal behaviours as psychological conditions, potentially leading to inappropriate treatments. So they laid out their concerns in an open letter, co-sponsored by five divisions of the American Psychological Association in Washington DC. “I thought, ‘Well, maybe we’ll get a couple or maybe 30 signatures,’” says Elkins, an emeritus professor at Pepperdine University in Malibu, California.

But the letter, posted online on 22 October (go.nature.com/uhmvmqq), touched a nerve. Within 10 days more than 2,800 people had signed it, many identifying themselves as mental-health professionals.

The petition targets proposed revisions to the *Diagnostic and Statistical Manual of Mental Disorders (DSM)*, a tome used by psychiatrists, psychologists, counsellors and others worldwide to diagnose mental maladies and set research agendas. The American Psychiatric Association, based in Arlington, Virginia,

plans to publish a new edition of the manual, *DSM-5*, in 2013. The association has declined to comment on Elkins’s petition.

Psychiatrist Allen Frances, who was the chief architect of *DSM-IV* and is an outspoken critic of its successor, has dubbed the open letter a “buyer’s revolt”. “I think the petition is the last best hope to influence the *DSM-5* from the outside,” says Frances, an emeritus professor at Duke University School of Medicine in Durham, North Carolina.

Elkins’s petition is not the first to raise concerns that the *DSM-5* proposals could overreach. In June, the British Psychological Society, based in Leicester, issued a critique that highlighted, for example, the proposed addition of ‘attenuated psychosis syndrome’. The society argued that this could be used “to stigmatize eccentric people”.

Elkins and his colleagues have complained about other proposals, such as the elimination of a ‘bereavement exclusion’ in the diagnosis of major depression. The previous edition of the

manual recommended that the condition not be diagnosed in people grieving the death of a loved one within the previous two months. The revisions shorten this to two weeks, a change that troubles psychiatrist Ramin Mojtabai of the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland. Categorizing these patients as having depression could boost the use of medications when psychotherapy may be the better treatment, he says.

Efforts to tighten loose definitions of attention deficit and hyperactivity disorder (ADHD) and bipolar disorder in children have also proved controversial. In response to worries that inexact criteria may have contributed to a surge in diagnoses of these conditions since the 1990s, the *DSM-5* task force has proposed a syndrome called ‘disruptive mood dysregulation disorder’, which would provide an alternative to labelling a child as bipolar or having ADHD. But Frances says that is not enough. “There should be a black box warning about how child bipolar disorder is being overdiagnosed,” he says. “Instead, they’ve created a new disorder.”

Field trials of the proposed *DSM-5* criteria have been completed and investigators plan to publish the results. Helena Kraemer, a statistician and emeritus professor at Stanford University School of Medicine in Palo Alto, California, who is on the *DSM-5* committee, says that results from trials of some criteria will indicate whether they generate more frequent diagnoses.

But Mojtabai cautions that trial results may not reflect what will happen when *DSM-5* is published. “Any trial is artificial,” he says. “The clinicians in these trials have intensive training, but people who will use this manual in clinical practice will not receive that level of instruction.” ■

“There should be a black box warning about how child bipolar disorder is being overdiagnosed.”

MISCONDUCT

Report finds massive fraud at Dutch universities

Investigation claims dozens of social-psychology papers contain faked data.

BY EWEN CALLAWAY

When colleagues called the work of Dutch psychologist Diederik Stapel too good to be true, they meant it as a compliment. But a preliminary investigative report (go.nature.com/tqmp5c) released on 31 October gives literal meaning to the phrase, detailing years of data manipulation and blatant fabrication by the prominent Tilburg University researcher.

"We have some 30 papers in peer-reviewed journals where we are actually sure that they are fake, and there are more to come," says Pim Levelt, chair of the committee that investigated Stapel's work at the university.

Stapel's eye-catching studies on aspects of social behaviour such as power and stereotyping garnered wide press coverage. For example, in a recent *Science* paper (which the investigation has not identified as fraudulent), Stapel reported that untidy environments encouraged discrimination (*Science* 332, 251–253; 2011).

"Somebody used the word 'wunderkind'," says Miles Hewstone, a social psychologist at the University of Oxford, UK. "He was one of the bright thrusting young stars of Dutch social psychology — highly published, highly cited, prize-winning, worked with lots of people, and very well thought of in the field."

In early September, however, Stapel was suspended from his position as dean of the Tilburg School of Social and Behavioral Sciences over suspicions of research fraud. In late August, three young researchers under Stapel's supervision had found irregularities in published data and notified the head of the social-psychology department, Marcel Zeelenberg. Levelt's committee joined up with sister



Dutch psychologist Diederik Stapel.

committees at the universities of Groningen and Amsterdam, where Stapel has also worked, to produce the report. They are now combing through his publications and their supporting data, and interviewing collaborators, to map out the full extent of the misconduct.

MISTAKES MADE

Stapel initially cooperated with the investigation by identifying fraudulent publications, but stopped because he said he was not physically or emotionally able to continue, says Levelt. In a statement, translated from Dutch, that is appended to the report, Stapel says: "I have made mistakes, but I was and am honestly concerned with the field of social psychology. I therefore regret the pain that I have caused others." *Nature* was unable to contact Stapel for comment.

The report does not identify specific papers

that contain manipulated or fabricated data, pending the completion of the investigations. The investigators conclude, though, that Stapel acted alone. "The co-authors, and in particular the PhD students, were absolutely not involved, they really didn't know what was going on in this data fabrication," Levelt says.

Often, the report says, Stapel and a colleague or student came up with a hypothesis, and then designed an experiment to test it. Stapel took responsibility for collecting data through what he said was a network of contacts at other institutions, and several weeks later produced a fictitious data file for his colleague to write up into a paper. On other occasions, Stapel received co-authorship after producing data he claimed to have collected previously that exactly matched the needs of a colleague working on a particular study.

The data were also suspicious, the report says: effects were large; missing data and outliers were rare; and hypotheses were rarely refuted. Journals publishing Stapel's papers did not question the omission of details about where the data came from. "We see that the scientific checks and balances process has failed at several levels," Levelt says.

At a press conference, Tilburg University's rector, Philip Eijlander, said that he would pursue criminal prosecution of Stapel. The committee is also producing a list of tainted papers to guide co-authors and journal publishers in what will probably be a long list of retractions.

Joris Lammers, a psychologist at Tilburg who did his PhD under Stapel's supervision, says he is "shocked" by the findings. Lammers says he worked independently of Stapel and collected all the data in his PhD himself — the report notes that his dissertation is not under suspicion. Several other former collaborators contacted by *Nature* declined to comment.

Hewstone, who has never worked with Stapel, had initially fretted that Stapel's fraudulent oeuvre would undermine other findings in the field of social psychology. While editing a new edition of a social-psychology textbook, however, Hewstone turned up no references to Stapel's work in 15 chapters, suggesting that Stapel's work was not as influential as he had thought. "I think the impact is going to be particularly devastating for the young people he worked with, but not for the field of social psychology as such," he says. ■

PERSBUREAU VAN ELJINDHOVEN



TOP STORY



Clues from China add to knowledge of genetics behind schizophrenia
go.nature.com/ia4cce

MORE NEWS

- Gaps in satellite coverage will spark data crisis go.nature.com/qkyc4z
- Transgenic rice makes human blood protein go.nature.com/e4rqkt
- Gulf universities hope for research funds go.nature.com/imzn7f
- Spanish institute faces cash crisis go.nature.com/wwntip



In April, the *Chikyu* will set out to drill through the fault that caused the Tohoku earthquake.

SEISMOLOGY

Drilling ship to probe Japanese quake zone

Fast-tracked expedition will measure fault's residual heat.

BY NICOLA JONES

After being tossed about and damaged by the tsunami that devastated northeastern Japan on 11 March, Japan's drilling ship the *Chikyu* has been given an especially fitting assignment: to drill into the fault zone and take temperature measurements near the epicentre of the magnitude-9.0 Tohoku earthquake that caused the tsunami. It will be the first time that researchers have drilled into an underwater fault soon after a quake. The aim of the exercise is to solve a decades-old mystery about the part that friction plays in such an event. This should help scientists to understand why some faults are more likely than others to cause tsunamis — in this case, one that ultimately claimed more than 23,000 lives.

"It would be a great disservice to society if we did not learn as much as possible from the fault zone heated by this huge earthquake," says Kiyoshi Suyehiro, president and chief executive of the management group of the Integrated Ocean Drilling Program (IODP). Following its initial approval of the proposal in September, the IODP has now confirmed that funding is available for the *Chikyu* to set sail in April and drill at a site south of the quake's epicentre (see map).

The scientific rationale for the expedition,

officially called the Japan Trench Fast Drilling Project, is detailed in a 2009 report promoting rapid-response drilling through fault lines as soon as possible after an earthquake in which the ground slips by more than one metre. The Tohoku event set a new record for the greatest amount of slippage ever observed — a whopping 50 metres — making it an ideal target.

"It's a fundamental issue in seismology right now: how do you get rock to slip tens of metres?" says James Mori, a seismologist at the Disaster Prevention Research Institute of Kyoto University in Japan, a co-author of the rapid-response report and joint chief of the upcoming drilling expedition. Researchers think that an important part of the answer is that resistance between the plates of rock, sand and water in

a fault line drops significantly during a quake — because of rock melting or increased water pressure, for example — but no one has been able to measure this effect properly. Because friction is dissipated as heat, precise temperature data should fill a crucial knowledge gap.

"We did a lot of planning, not knowing what kind of quake we'd have to do it with," says Emily Brodsky of the University of California, Santa Cruz, who is also involved in the mission.

Researchers have attempted to monitor the underground temperature after an on-land earthquake on three previous occasions — after the 1995 Kobe earthquake in Japan, the 1999 Chi Chi quake in Taiwan and the 2008 Wenchuan quake in Sichuan Province, China. But these projects produced only a few temperature readings between them, and found only tiny temperature increases, or nothing at all — perhaps because the temperature rise was too small to see, or because of imperfect monitoring techniques. "The recurring theme is that the faults tend to be colder than they should be," says Brodsky. A larger slip event provides a better chance of tracking the expected temperature increase of up to 0.5 °C, she says. "We need to do this now, and do it fast, and do it correctly."

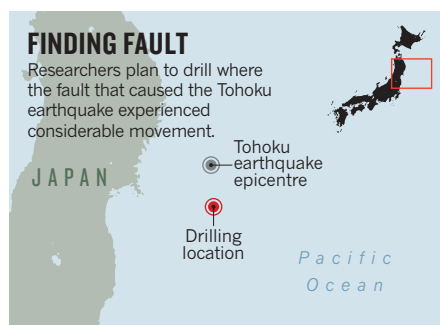
The *Chikyu* will drill down 1 kilometre through the fault, and drop a string of temperature sensors down the hole. By tracking temperatures for one to three years — much longer than has been attempted before — researchers should be able to calculate the total amount of heat that was generated by the quake. That will provide them with the resistance forces felt in the fault during the slip, filling in a blank in models of earthquake dynamics. "This is a key missing ingredient," says Jean-Philippe Avouac, a geologist at the California Institute of Technology in Pasadena, who is not involved in the project.

Completing the drilling won't be easy. At the proposed site, the Tohoku fault lies under 7 kilometres of water and some 700 metres of Earth's crust, so a huge drill string will be needed. Previously, only a tiny 15-metre core has ever been extracted from beneath water of that depth, says Brodsky; most cores are taken from beneath 6 kilometres of water or less.

In addition to temperature measurements, the project will also examine the sediments pulled up in the core. Certain sediment textures, such as ball-bearing-like particles of clay, might be associated with large-slip earthquakes. Identifying such features should help scientists to forecast the slip potential of other faults.

The chance to collect precious information from the Tohoku event represents "an opportunity, maybe even a responsibility," says Mori. Almost all of the damage caused by the quake was done by the tsunami, he points out.

"What we really want to understand is what caused that." ■



➔ **NATURE.COM**
For more on ocean floor faults see:
go.nature.com/6z7bwq

GENOMICS

Research cuts hit the DNA business

Makers of sequencing machines hope for profits in the clinic.

BY ERIKA CHECK HAYDEN

A DNA sequencing machine is one of the most expensive lab purchases a biologist can make. So it is no surprise that, with research funding falling and worse times ahead, institutions are holding off buying the devices, leaving manufacturers feeling the pinch. After a round of gloomy third-quarter results, many makers of sequencing machines are now hoping that cost cutting and expansion into medical diagnostics will help them to prosper again.

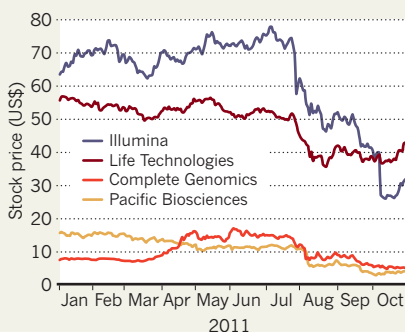
Businesses that make sequencing machines or offer sequencing services rely heavily on custom from federally funded researchers, particularly those supported by the US National Institutes of Health (NIH). Although the NIH's 2011 budget is only 1% lower than it was in 2010, pessimism about future government spending has led analysts to predict much bigger budget cuts in 2012 and 2013 — and that pessimism has also had an impact on sequencing companies' share prices (see 'Taking a hit').

"Owing to fears of lower government funding, institutions are certainly being more conservative with expenditures," says Shawn Levy, a geneticist at the HudsonAlpha Institute for Biotechnology, a genomics research centre in Huntsville, Alabama. Indeed, in a survey released on 24 October, 58% of 94 genome scientists polled by the online publication GenomeWeb and the Tokyo investment bank Mizuho Securities said that talk of austerity had caused them to delay lab purchases. Forty-two per cent said that they would probably not buy a new DNA sequencing machine in the next 12 months.

The bleak mood was clear in a round of financial results announced last week. Illumina of San Diego, California, which has captured the largest share of the sequencing market, said that its third-quarter profits were down by 43% compared with the same quarter of 2010. The company blamed high costs and poorer-than-expected sales for the decline, and said that it will lay off some 200 workers, about 8% of its global workforce. Life Technologies of Carlsbad, California, said that its revenue grew by 7% in the last quarter, compared with the same period in 2010, helped

TAKING A HIT

Manufacturers of sequencing machines saw stocks fall in July in response to planned US spending cuts.



by strong sales of its Ion Torrent machines. But with an eye to difficult times ahead, the company has already laid off workers and carried out a round of cost cutting.

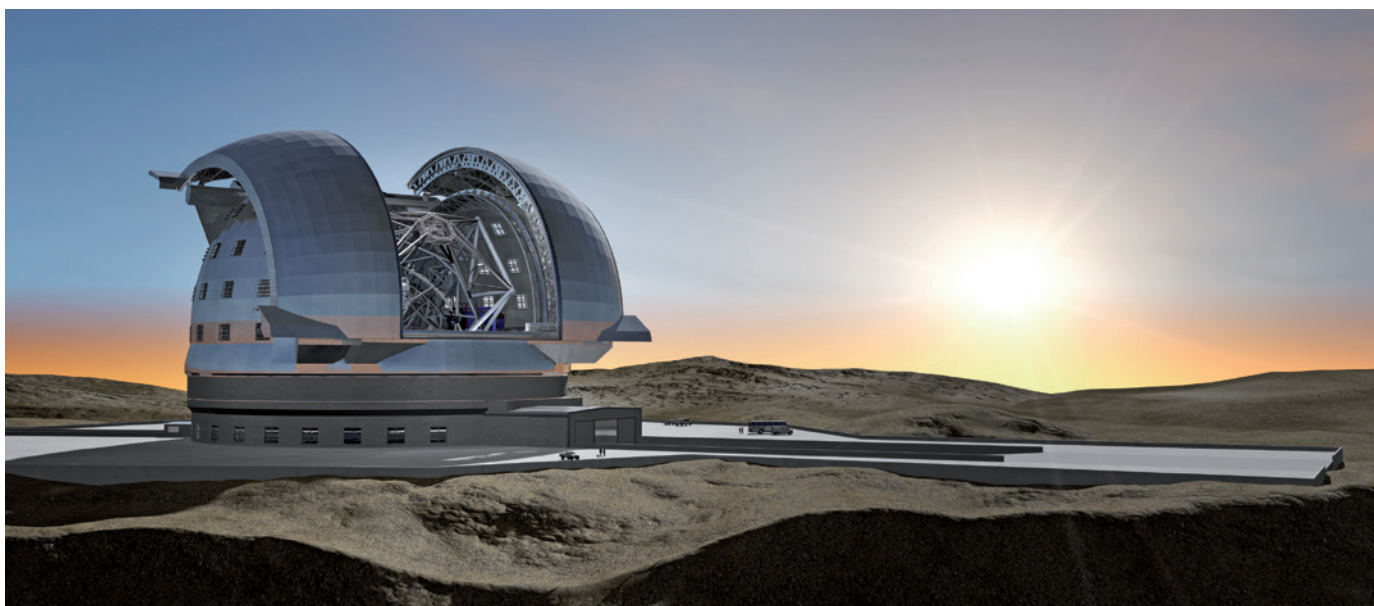
In September, Pacific Biosciences of Menlo Park, California, was forced to lay off 130 people — 28% of its staff — and reduce its overall 2011 earnings forecast. Other major genomics companies, such as California-based Complete Genomics of Mountain View and Affymetrix of Santa Clara, reveal their earnings this month.

In light of the dismal news, many sequencing companies are now eyeing new markets in biomedicine, where whole-genome sequencing is beginning to make inroads (see *Nature* 478, 22–24; 2011). Illumina, for example, will set up a business unit to promote the use of its tools in clinical diagnostics, and hopes that the imminent release of a lower-cost version of its MiSeq sequencing machine will boost sales.

Life Technologies, meanwhile, boasted that it would develop a genetic test to help predict an individual's response to a cancer drug that UK-based drug-maker GlaxoSmithKline is studying in phase III clinical trials. It also touted a US\$9-million deal to sell forensic-analysis systems in Russia, and underscored its double-digit growth in China, saying that it will increasingly turn its attention to Asia.

Dan Leonard, an analyst at health-care investment consultants Leerink Swann in Boston, Massachusetts, agrees that moving into the clinic is ultimately the sequencing industry's best hope for growth. ■

➔ **NATURE.COM**
Nature's genomics
supplement:
go.nature.com/ftkwlr



ESO

The European Extremely Large Telescope in Chile, as envisioned when completed around 2022, may be the first instrument to witness the birth of galaxies.

ASTRONOMY

Megatelescopes look for support

As a huge European observatory nears approval, a National Science Foundation funding competition between equivalent US projects remains in a holding pattern.

BY ERIC HAND

Astronomers in the United States may have to do without government support in the race to create the world's biggest telescopes and gather photons from the Universe's first stars. Two projects — the Thirty Meter Telescope (TMT) and the Giant Magellan Telescope (GMT) — are vying for support from the US National Science Foundation (NSF), but the cash-strapped agency says that it may not be able to fund either before 2020, by which time both projects had hoped to be finished (see *Nature* **469**, 451; 2011). That makes it even more likely that a comparable project, the European Extremely Large Telescope (E-ELT), will be ready years ahead of its US-led counterparts.

Jim Ulvestad, director of the NSF's astronomy division, is charged with holding a competition to choose which project the agency should support — a move emphatically recommended by the 2010 decadal survey in astronomy, an assessment of priorities for US agency planners. "But the decadal survey also assumed we'd have a lot more money than we do," Ulvestad says.

The problem, says Ulvestad, is a long queue of major new facilities that the NSF is already committed to funding between now and the end of the decade (see 'Join the queue'). Even the cancellation of the Deep Underground Science and Engineering Laboratory in December

2010 (see *Nature* <http://dx.doi.org/10.1038/news.2010.683>; 2010) is not enough to overcome a poor allocation for such projects from Congress earlier this year.

Ulvestad says the agency may still launch a competition in December or January — although the winner would not be guaranteed any money. The competition could coincide with approval of the 39-metre E-ELT, planned for Cerro Armazones in Chile. At a council meeting in December, member states of the European Southern Observatory (ESO), based in Garching, Germany, plan to authorize limited construction on the €1.1-billion (US\$1.6-billion) telescope — a move that could lure frustrated international partners away from the two US-led efforts.

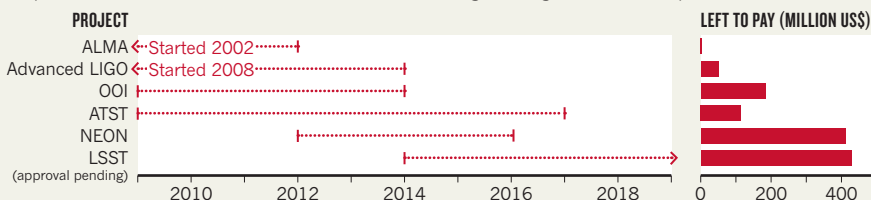
The US teams are at odds in their views on an

NSF competition. The consortium behind the \$1-billion TMT, planned for the top of Mauna Kea in Hawaii, thinks it has a good chance of besting its rival; it has lobbied for years for the NSF to pick a winner. "Let's get on with it," says Richard Ellis, a member of the TMT board and an astronomer at the California Institute of Technology in Pasadena. "Let's at least make plans, even if the money comes later."

Ellis says that NSF backing would help the TMT to nail down other partner nations, especially China, which wants to see US government action before it commits completely. He adds that even if the NSF lacks money for construction, it could become a partner by paying operating expenses later on. And holding a competition might compel the agency to find money in the future. "The NSF has to rise to

JOIN THE QUEUE

Commitments to a succession of projects (left) with large capital costs (right) over the next 10 years may keep the US National Science Foundation from contributing to a large US-led telescope until 2020.



ALMA, Atacama Large Millimeter/submillimeter Array; LIGO, Laser Interferometer Gravitational Wave Observatory; OOI, Ocean Observatories Initiative; ATST, Advanced Technology Solar Telescope; NEON, National Ecological Observatory Network; LSST, Large Synoptic Survey Telescope.

the challenge of lobbying for facilities,” says Ellis. “It has been too passive.”

The GMT prefers the passive approach. The 25-metre telescope, planned for Las Campanas, Chile, is the smallest of the projects, and will be about \$300 million cheaper than the TMT. The involvement of wealthy institutions such as Harvard University in Cambridge, Massachusetts, is assured, so the consortium might be able to raise its capital without the NSF. It already has something to show for its efforts: astronomers cast the first of the GMT’s seven mirrors in 2005, and since then have been configuring it to a unique off-axis reflecting surface. Wendy Freedman, chairwoman of the GMT board and director of the Carnegie Observatories in Pasadena, says that the GMT still wants to partner with the NSF, but would prefer to hold off on a competition.

“Making decisions without money strikes us as an unusual thing to do,” she says. “The logical thing to do would be to wait.” She doesn’t agree that NSF backing will help to attract partners; the GMT already has financial commitments from institutions in Australia and South Korea. Ultimately, says Freedman, the GMT may elect not to compete.

The ESO has a steady stream of funding, but

NATURE.COM
For more on the
megatelescopes,
visit:
go.nature.com/hahede

even so, the E-ELT’s price tag could prove problematic. Current treaty-enforced annual fees from the 14 member states will account for only about one-third of the cost. Brazil became the source for another third when it agreed in December 2010 to join the ESO. And on 13 October, Chile formally donated the site on which the E-ELT will be built.

But the observatory still has to find the remaining third of the cost, says Tim de Zeeuw, director-general of the ESO. One way would be to recruit yet more members. De Zeeuw notes that a number of countries are “asking questions” about membership: Russia, Poland, Canada (currently a TMT member) and Australia (committed to the GMT). But in the meantime, de Zeeuw is trying to muster approval from the member states for an annual 2% fee increase, along with a one-time special contribution proportional to the state’s income.

Three nations — Sweden, Finland and the Czech Republic — have agreed to unlock the money, he says. But the remaining 11 are not likely to rally approval by December; nor is the Brazilian government expected to have ratified its membership treaty by then.

Approval of the full E-ELT budget won’t come before a meeting in March 2012. Instead, the council is planning to approve early contracts and crucial infrastructure, such as a road to the E-ELT’s mountain location. But

even this limited construction authority will give the project momentum. “They could have said, ‘We’ll decide next year,’” says Roberto Gilmozzi, principal investigator of the project.

The US competition, once begun, would probably be decided in less than a year. Ulvestad says that the NSF would want to evaluate the risk of the projects, and would ask how, for instance, the telescope consortia might provide publicly accessible archives or user support for astronomers not associated with consortium institutions.

But Ulvestad also realizes that, with so long until the money will materialize, the telescopes might be completed without the NSF’s help. Much of the US academic astronomy community relies on NSF support to get access to telescope time, which might not be possible with the giant telescopes unless the NSF has contributed to them. But, he says, with a total of seven US universities signed on to the TMT or GMT, there is already “a lot of US community in their partnerships”. ■

CORRECTION

The News story ‘Malaria vaccine results face scrutiny’ (*Nature* **478**, 439–440; 2011) wrongly described *Plasmodium falciparum* as multicellular. The parasite is unicellular.



The autism enigma

Diagnoses and research funding are rising, but much about autism remains a puzzle. *Nature* seeks some truths.

Everything about autism spectrum disorder conspires to make it hard to understand. It takes diverse forms, from profound communication and behavioural problems to social difficulties coupled with normal language and even precocious talents. (Here, *Nature* will refer to them all as autism.) The prevalence of autism is rising — by some counts, steeply — but the reasons for that are unclear. Causes of the condition include a complicated mixture of genetic and environmental factors, most unknown (see page 5). Its roots lie in the development of the human brain, a process that, despite huge leaps in neuroscience, remains mysterious. So as awareness rises and parents clamour for answers, scientists

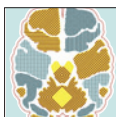
can offer few certainties. Hearsay and unsubstantiated theories sometimes fill the void.

This week, *Nature* searches for some truths about autism. Some researchers have evidence to combat the notion that the rise in prevalence can all be explained by changes in how the condition is diagnosed (see page 22). Others are debating the idea that some scientists and engineers are themselves 'on the spectrum', and are at high risk of having a child with autism (see page 25). At the same time, researchers are learning that although autism is clearly a disability, certain characteristics of it could be an advantage in science

(see page 33). A debunked link between vaccines and autism still clouds the public discussion, but some advocates have taken a firm stand in favour of rigorous science, and the answers it will eventually provide (see page 28). Much more content can be found at nature.com/autism.

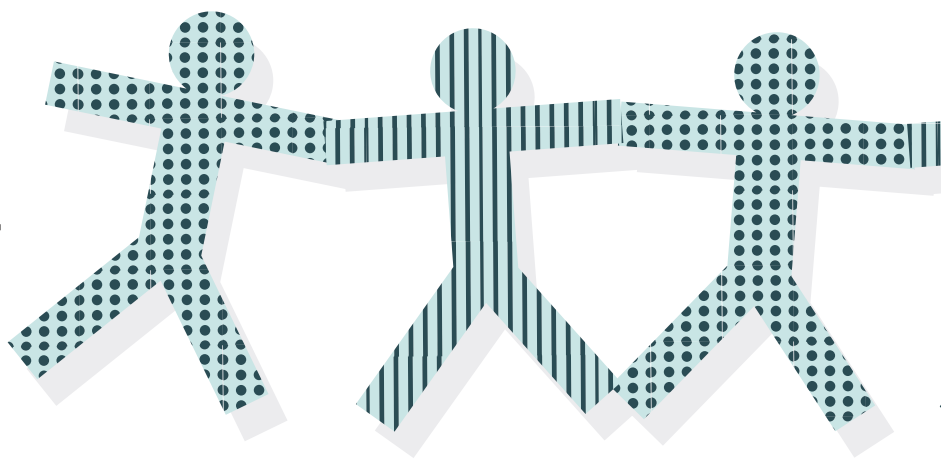
Even before fundamental problems are solved, research is revealing better ways to support people with autism. If the condition is diagnosed early, a growing repertoire of evidence-based therapies can be applied to give children the best possible chance of living full lives. Meanwhile, the spotlight on autism is helping to reduce stigma.

The complexities that make autism hard to understand are a magnet for researchers — and this should lead to a future with less fiction and more much-needed fact. ■



THE AUTISM ENIGMA
Sorting fact from fiction
nature.com/autism

Autism counts



Shifting diagnoses and heightened awareness explain only part of the apparent rise in autism. Scientists are struggling to explain the rest.

BY KAREN WEINTRAUB

When Leo Kanner first described autism in 1943, he based his observations on 11 children with severe communication problems, repetitive behaviours such as rocking and an acute lack of social interaction. The physician and psychiatrist at Johns Hopkins University in Baltimore, Maryland, predicted that there were probably many more cases than he or anyone else had noticed¹. “These characteristics form a unique ‘syndrome’, not heretofore reported,” he wrote, “which seems to be rare enough, yet is probably more frequent than is indicated by the paucity of observed cases.”

Kanner’s prophecy has been more than fulfilled. An early study², in 1966, examined eight- to ten-year-old schoolchildren in Middlessex, UK, and estimated a prevalence of 4.5 cases per 10,000 children. By 1992, 19 in every 10,000 six-year-old Americans were being diagnosed as autistic³.

Numbers skyrocketed in the first decade of the twenty-first century, according to data from the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia. Surveying what is now known as autism spectrum disorder (ASD), the CDC found that by 2006, more than 90 in 10,000 eight-year-olds in the United States had autism⁴. Put another way, autism was now affecting 1 in every 110 children — a figure that strengthened public fears that an ‘epidemic’ was afoot (see ‘Diagnosis: rising’).

For the most part, research into autism’s prevalence had explained away the increase. Studies attributed it to greater awareness of

the condition, the wider diagnostic criteria for ASD, more frequent diagnosis of children with mental retardation as also having autism and diagnosis at younger ages. But by the mid-2000s, researchers started to note that these explanations were coming up short. “A true risk due to some, as yet to be identified, environmental risk factor cannot be ruled out”, read one study from 2005 (ref. 5).

That shift is important. If the rise in autism can be explained mainly by increased awareness, diagnosis and social factors, then the contributing environmental factors will always have been present — perhaps an ill-timed infection in pregnancy or some kind of nutritional deficit. If the increase can’t be explained away — and at least part of the rise is ‘real’ — then new factors must be causing it, and scientists urgently need to find them.

The subject is sensitive. Parents of children with autism agonize over whether they could have done something to prevent it. Researchers have been wary of invoking environmental triggers because that harkens back to a long-discarded idea that cold, unloving ‘refrigerator’ mothers were the source of their children’s problems. And the increase in prevalence has been used to support more recently debunked hypotheses such as the idea that vaccines cause autism.

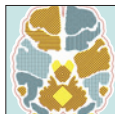
Thomas Insel, director of the National Institute of Mental Health in Bethesda, Maryland, says it is time to get past these

legacies. “This whole idea of whether the prevalence is increasing is so contentious for autism, but not for asthma, type 1 diabetes, food allergies — lots of other areas where people kind of accept the fact that there are more kids affected.” To him, it is clear that there is a real increase in autism, and researchers need more funding and encouragement to look at possible environmental causes. During the past decade, the US federal government has spent about US\$1 billion researching the genetics of autism and only about \$40 million on studies of possible environmental factors.

Not everyone agrees with Insel’s assessment. Some argue that the current data aren’t strong enough to say for certain that the increase in autism diagnoses represents a true change in its prevalence. “It feels like the numbers are going up. It really does,” says Richard Grinker, an anthropologist at George Washington University in Washington DC. But “when I look at the science, that doesn’t stand up”, he says. “You simply can’t take prevalence estimates of autism as if they are the kind of hard scientific evidence that you would get from mapping out the increase in a virus.”

CHANGING CRITERIA

No one knows for sure what causes autism, although genes and environment both seem to be involved. The brain’s white matter may grow too fast in the first two years of life, leaving its networks jumbled. Synapses, the junctions between neurons, might not be functioning normally. Or other physiological processes could be involved: autism has



THE AUTISM ENIGMA

Sorting fact from fiction
nature.com/autism



been variously linked to epilepsy, digestive problems, immune or hormonal dysfunction, mitochondrial function and more.

The diagnostic criteria for autism have changed over time. In 1952, autism defined by Kanner's narrow description was diagnosed as 'early-onset schizophrenia'; it was renamed 'infantile autism' in 1980 and then 'autism disorder' in 1987. In the past decade, the common name autism has covered a wider range of behavioural, communication and social disorders also referred to by the umbrella term ASD, which includes autistic disorder, Asperger's syndrome and other related conditions.

Diagnoses of autism are also subjective. Social skills vary widely in the general population, as do other behaviours associated with autism. When does lack of spontaneity or an inability to make eye contact become a problem worthy of a medical label? And the frequency of diagnosis often reflects how eager parents are to receive one. When there's a stigma attached, diagnoses are likely to fall; when public support rises, so will cases.

A diagnosis is mutable, says Grinker. "It is a framework for a set of symptoms. And it's a framework that works at a particular point in time with a certain society and a certain health-care system and education system, and that will change as society changes."

Such considerations help to explain the startlingly high prevalence of autism that Grinker found in South Korea in a study published this year⁶. In the 1980s, he had found Korean families generally unwilling to admit that anything might be wrong with their children, because of the stigma attached⁷. But when he undertook the latest study, attitudes had changed. Families in Ilsan, a stable, residential community on the outskirts of Seoul, welcomed information about autism, which in this study was offered confidentially. His team screened more than 55,000 children born between 1993 and 1999, and came up with an estimated prevalence for ASD of 1 in 38 (ref. 6). Grinker says that this is perhaps an

overestimate, but it's the best his team could produce.

Current US prevalence figures for autism are likely to be too low, Grinker says, because they don't look at the entire population. Many US studies are based on diagnosed cases of autism, either in the California school district — the nation's largest — or in the CDC's Autism and Developmental Disabilities Monitoring Network. But the California data count only children old enough to be in school and disabled enough to get a diagnosis or need services. The CDC surveillance also only picks up children with a documented developmental disorder. These methods probably miss children at the milder end of the spectrum.

Some research suggests that the prevalence has always been high. In a study published this year⁸, a team led by Terry Brugha, a psychiatrist at the University of Leicester, UK,

counting cases of autism in the 1970s. He found a prevalence of autism of 0.7% among seven-year-old Swedish children in 1983 (ref. 9) and 1% in 1999 (ref. 10). "I've always felt that this hype about it being an epidemic is not really very likely," he says.

FILLING THE GAP

Nevertheless, with numbers rising fast, many expect to see some sort of smoking gun in the environment. Peter Bearman, a sociologist at Columbia University in New York, has been trying to figure out how much of the increase is driven by social forces. He analysed nearly 5 million California birth records and 20,000 records from the state's department of developmental services. By linking birth with detailed diagnostic data he was able to generate a rich picture of the demographics and life history of those with autism, which yielded clues to the social factors that influence diagnosis.

"This whole idea of whether the prevalence is increasing is so contentious for autism."

looked into autism's past by counting adults with the disorder. His team knocked on more than 7,000 doors across England. And although Brugha expected to uncover a very low prevalence of autism in adults, he and his colleagues calculated it as 9.8 in 1,000 — close to the frequency found in US children.

Brugha says that the research needs to be repeated in different groups, but the implication is that autism prevalence is stable. "If this is confirmed in other studies, it means we should also be looking for causes of autism that have always been there, and not just for causes that have developed in recent years or decades," he says.

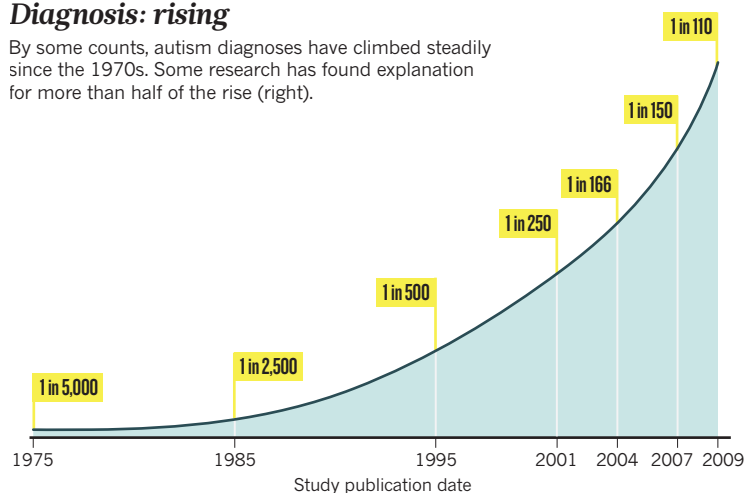
Christopher Gillberg, who studies child and adolescent psychiatry at the University of Gothenburg in Sweden, has been finding much the same thing since he first started

So far, Bearman says, he can account for just more than 50% of the observed increase (see 'Reasons: unclear'). Around 25% of the rise in autism over the past two decades can be attributed to what he calls "diagnostic accretion". He could see from the medical records that some children who would have been diagnosed as mentally retarded a decade ago are now given a diagnosis of both mental retardation and autism¹¹. Another 15% can be accounted for by the growing awareness of autism — more parents and paediatricians know about it¹².

Geographic clustering accounts for 4%, Bearman says. The most fascinating cluster lies in and around the hills of Hollywood, California. Children living in a 900-square-kilometre area centred on West Hollywood are four times more likely to be diagnosed with autism than are those living elsewhere

Diagnosis: rising

By some counts, autism diagnoses have climbed steadily since the 1970s. Some research has found explanation for more than half of the rise (right).



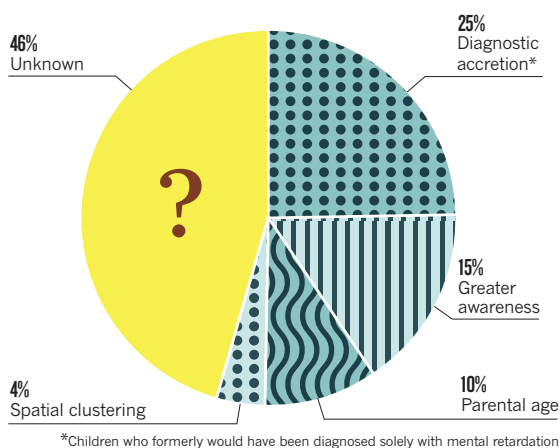
in the state¹². Some residents worried that something in the water was triggering autism — perhaps the legacy of a 1959 nuclear accident at the Santa Susana Field Laboratory in nearby Simi Valley — but Hollywood shares its water supply with Los Angeles, where autism rates are not uniformly high. Moreover, rates are high whether families have lived in Hollywood for years or have just moved there, Bearman says.

He suspects that the real reason for the cluster has to do with neighbourliness: a parent explains to a neighbour over the back fence where to find help and how to navigate the medical and educational systems. Once a cluster of informed, involved parents builds up, specialists are more likely to settle in that

something in the environment. Rather than quibbling over recounts they are focusing on finding the causes.

Since autism was first identified, ideas about its cause have swung to and fro between nature and nurture. The early focus on ‘refrigerator’ mothers resulted in a backlash and a stronger focus on genetics. The pendulum now seems to have settled somewhere in the middle, which is where many think it should be. “The bulk of the autism research that’s occurred has only looked at genetics,” says Lisa Croen, director of the autism research programme at the health-insurance provider Kaiser Permanente, in Oakland, California. “We’ve learned a lot but we haven’t found the magic bullet. I think that’s because part of the

Reasons: unclear



*Children who formerly would have been diagnosed solely with mental retardation

“If it is an environmental cause contributing to an increase, we certainly want to find it.”

area, diagnosing and treating even more kids, Bearman says.

Another 10% of the increase may be explained by a social change with biological implications: people having children when they are older. Some research has found that children born to parents older than 35 have a higher risk of being diagnosed with autism. Studies are divided about whether the mother’s age or the father’s has the strongest influence, but Bearman’s work on parents older than 40 suggests that the mother’s age matters more¹³.

The fact that he still cannot explain 46% of the increase in autism doesn’t mean that this ‘extra’ must be caused by new environmental pollutants, Bearman says. He just hasn’t come up with a solid explanation yet. “There are lots of things that could be driving that in addition to the things we’ve identified,” he says.

But many researchers now say that at least part of the rise in autism is real and caused by

picture has been missing.”

Several major federally funded trials, together with other smaller ones, are now under way in the United States and elsewhere to try to fish out what makes a child autistic. They are hoping to uncover unknown risk factors and markers for autism by monitoring environmental exposures and taking regular biological samples from children and their parents.

In 2007, for example, the Study to Explore Early Development (SEED), under the auspices of the CDC, began recruiting about 2,700 children aged two to five. The study includes developmental evaluations, questionnaires, a review of medical records and analysis of blood, cheek-cell and hair samples to examine genetic make-up and exposures to environmental chemicals. The Early Autism Risk Longitudinal Investigation (EARLI), funded by the National Institutes of Health, is enrolling up to 1,200 families that have a

child with autism and are preparing to have another baby. The study intends to look for any interplay between environmental factors and genetic susceptibility that might contribute to autism risk in their next child.

“These studies are really going to fundamentally change the landscape,” says Croen, who is a lead investigator on SEED. She and others expect a dramatic improvement in the understanding of autism and its prevalence over the next five to ten years.

Craig Newschaffer, an epidemiologist at Drexel University in Philadelphia, Pennsylvania and an investigator with EARLI, says that a focus on the rise in diagnoses may be less important than figuring out what is causing autism in the first place. “If it is an environmental cause that’s contributing to an increase,” he says, “we certainly want to find it.” It may be time to move on from the question of whether or not autism is truly rising, “I think it’s probably a nearly intractable question to answer.” ■

Karen Weintraub is a freelance writer in Cambridge, Massachusetts.

1. Kanner, L. *Nerv. Child* **2**, 217–250 (1943).
2. Lotter, V. *Soc. Psychiatry* **1**, 124–137 (1966).
3. Newschaffer, C. J., Falb, M. D. & Gurney, J. D. *Pediatrics* **115**, e277–e282 (2005).
4. ADDMN Surveillance, Year 2006, Principal Investigators *MMWR Surveill. Summ.* **58**, 1–20 (2009).
5. Rutter, M. *Acta Paediatr.* **94**, 2–15 (2005).
6. Kim, Y. S. et al. *Am. J. Psychiatry* **168**, 904–912 (2011).
7. Grinker, R. R. *Korea and its Futures: Unification and the Unfinished War* (St Martins Press, 1998).
8. Brugha, T. S. et al. *Arch. Gen. Psychiatry* **68**, 459–465 (2011).
9. Gillberg, C. *J. Child Psychol. Psychiatry* **24**, 377–403 (1983).
10. Kadesjö, B., Gillberg, C. & Hagberg, B. *J. Autism Dev. Disord.* **29**, 327–331 (1999).
11. King, M. & Bearman, P. *Int. J. Epidemiol.* **38**, 1224–1234 (2009).
12. King, M. D. & Bearman, P. *S. Am. Sociol. Rev.* **76**, 320–346 (2011).
13. King, M. D., Fountain, C., Dakhllallah, D. & Bearman, P. *S. Am. J. Public Health* **99**, 1673–1679 (2009).



PETE ELLIS/DRAWGOOD.COM

When geeks meet

Psychologist Simon Baron-Cohen thinks scientists and engineers could be more likely to have a child with autism. Some researchers say the proof isn't there.

In the opening scene of *The Social Network*, Jesse Eisenberg portrays a cold Mark Zuckerberg getting dumped by his girlfriend, who is exasperated by the future Facebook founder's socially oblivious and obsessive personality. Eisenberg's Zuckerberg is the stereotypical Silicon Valley geek — brilliant with technology, pathologically bereft of social graces. Or, in the parlance of the Valley: 'on the spectrum'.

Few scientists think that the leaders of the tech world actually have an autism spectrum disorder (ASD), which can range from the profound social, language and behavioural problems that are characteristic of autistic disorder, to the milder Asperger's syndrome. But according to an idea that is creeping into the popular psyche, they and many others in professions such as science and engineering may display some of the characteristics of autism, and have an increased risk of having children with the full-blown disorder.

The roots of this idea can largely be traced

BY LIZZIE BUCHEN

to psychologist Simon Baron-Cohen at the University of Cambridge, UK. According to a theory he has been building over the past 15 years, the parents of autistic children, and the children themselves, have an aptitude for understanding and analysing predictable, rule-based systems — think machines, mathematics or computer programs. And the genes that endow parents with minds suited to technical tasks, he hypothesizes, could lead to autism when passed on to their children, especially when combined with a dose of similar genes from a like-minded mate¹.

The notion has an intuitive plausibility. In the public mind, it meshes with the stereotype of the scientist or computer geek as smart but socially awkward.

(Baron-Cohen has speculated that luminaries such as Albert Einstein and Isaac Newton had Asperger's syndrome.) And in scientific circles, many accept that certain autistic traits — social difficulties, narrow interests, problems with communication — form a continuum across the general population, with autism at one extreme. As most experts believe that genes have an important role in autism, it's also plausible that two parents with milder, 'autistic-like' traits could be more likely to have a child with autism.

It also fits at least some clinicians' experiences. "I see deep geeks of all sorts," says Bryna Siegel, a clinical psychologist who runs the autism clinic at the University of California, San Francisco, referring to the parents of children with autism. "They don't make great eye contact, all their clothing is from the Intel shop, they don't have a lot of social understanding. I do think that when these geeks marry each other, that's bad news for the offspring." But



THE AUTISM ENIGMA

Sorting fact from fiction
nature.com/autism

critics of Baron-Cohen's theories aren't hard to find. Autism researchers say that his work has focused primarily on a subset of people with 'high-functioning' autism — such as Asperger's syndrome — who have good language capabilities and at-least average intelligence. They say that the data are insufficient to support his theories and that many experiments cry out for independent replication.

"They're some really good hypotheses to think about, but they need to be tested," says John Constantino, a psychiatrist at Washington University in St Louis. "There's not a lot of data." Some critics are also rankled by Baron-Cohen's history of headline-grabbing theories — particularly one that autism is an 'extreme male' brain state. They worry that his theory about technically minded parents may be giving the public wrong ideas, including the impression that autism is linked to being a 'geek'.

Baron-Cohen acknowledges that "there is a problem that there are too few attempts at replication" of his studies, and says that he remains "open minded about these hypotheses until there are sufficient data to evaluate them". But he says he doesn't see a problem with introducing theories before definitive evidence has been collected. "I would see it as a positive contribution rather than a concern that scientists move from preliminary evidence to formulate the more general theory, especially when the theory is highly testable, since this is how science advances," he says.

BUCKING THE SYSTEM

In the 1990s, while most research on autism focused on problems with social interaction, Baron-Cohen became fascinated by the obsessive, narrow interests and repetitive behaviours that also characterize the condition. He noticed that children with autism were drawn to things such as machines, numbers, calendars and spinning objects². One child might memorize the technical specifications of gadgets; another would flip light switches on and off incessantly.

"The old view was that [such behaviours] were lacking in purpose, they just did it," says Baron-Cohen. But he started seeing these eccentricities from a new perspective. "They're figuring out how the family DVD player works, or understanding the electrical circuitry of the house. The child is doing it to understand the system." In autism, he theorizes, the brain has an average or superior ability to understand predictable systems, or 'hypersystemize', coupled with an inability to empathize, or understand other people's intentions and feelings.

Baron-Cohen cites several lines of evidence in support of his theory. In a 2003 study³, for example, he found that people with autism scored highly on the 'systemizing quotient', a questionnaire he designed. In a survey of undergraduates at the University of Cambridge, he found that those studying mathematics were more likely to have been



Simon Baron-Cohen noticed that children with autism are drawn to 'systems' such as machines and numbers.

diagnosed with autism than were students majoring in medicine, law or social science⁴. And, using another questionnaire called the autism quotient, he found that students in science and maths had higher scores on measures of autistic traits than did students in the humanities and social sciences⁵. Baron-Cohen says that although these surveys do not measure systemizing ability directly, they demonstrate that systemizing is a trait of autism, and also part of the 'broader autistic phenotype' that includes some of the wider population.

Baron-Cohen's critics, however, are sceptical of these surveys, in which subjects answer questions about themselves such as: 'I notice patterns in things all the time' and 'I would rather go to a library than to a party'. "Whether those self-perceptions, as with any of our self-perceptions, are accurate is questionable," says Francesca Happé, a cognitive neuroscientist at King's College London.

It would be more objective, say Happé and others, to test children with and without autism on their abilities to understand systems, and then compare the scores. "Rigorous studies are still missing," says Uta Frith, a developmental psychologist at University College London. "At the moment, he has people saying, 'yes, I'm a person interested in details', as opposed to actually observing them on tasks."

Baron-Cohen says that his lab is doing such follow-up work. He says that questionnaires can be advantageous because data can be collected quickly, and that even though biases can creep in, "you do find consistent patterns". He also points to a 2001 study⁶ in which he

showed that children with Asperger's syndrome can outperform typical children at figuring out how simple mechanical systems work. But critics counter that the children with Asperger's were selected on the basis of having average or above average IQs, whereas the typical children were selected at random. Similarly, critics point out that the Cambridge students with autism are highly unusual because they function well enough to attend one of the top universities in the world.

This is a common complaint about Baron-Cohen's work. "He's tended to focus on very bright individuals with ASD," says Catherine Lord, a clinical psychologist and autism researcher at Weill Cornell Medical College in New York. "A lot of the things he might say in describing those individuals are pretty irrelevant for most people with ASD."

Baron-Cohen acknowledges that "some of the psychological research is focused on high-functioning children with autism", because, he says, they have the language capabilities to perform the tests. "But my thought is that it could apply across the system," he says of the systemizing theory, to all children who have some form of the disorder.

Earlier this year, Liz Pellicano, a developmental psychologist at the Institute of Education in London, tested how a group of children with a wider range of ASD compare with a control group in figuring out a system. Her team designed a small room in which the floor was arrayed with 16 identical green lights. The children were asked to find the one light that, when pressed, would turn from green to red.

BLAZE ALBEMARLE/EYEVINE

The target light was on the same side of the room 80% of the time. Children with autism, including Asperger's syndrome, were much worse at figuring out this system than the children in the control group⁷. "They weren't systematic," says Pellicano. "When they were searching, they were unbelievably haphazard." In her view, she says, studies such as this show that Baron-Cohen's theory "isn't standing up to empirical tests".

Baron-Cohen says he is not sure that Pellicano's paradigm was testing the same sort of systemizing he describes. But, he says, he's "glad that at least people are starting to look at systemizing". So far, most work on the subject has come out of his lab. "I think our published studies are rigorous, but there are still too few studies into systemizing," he says. "It is still way too early to be able to look across dozens or hundreds of studies to evaluate that theory."

LIKE FATHER, LIKE SON?

Baron-Cohen proposes that systemizing ability can be inherited — and that in information-technology (IT) enclaves such as Silicon Valley, where hypersystemizers are more likely to meet, pair off and have children, the result is a higher incidence of autism.

Back in 1997, for example, he concluded that fathers of children with autism were more than twice as likely to be engineers as were fathers of non-autistic children⁸. But autism researchers Christopher Jarrold and David Routh at the University of Bristol, UK, pointed out that Baron-Cohen reported the analysis of data only for engineers, not for the other occupations surveyed. After analysing the same data⁹, they found that fathers of children with autism were more likely to work in medicine, science and accountancy, as well as engineering, and less likely to have manual occupations. They suggested that these fathers were simply more likely to have reached a higher level of education.

similar size¹⁰ — evidence that he takes as support for the idea that parents who are strong systemizers could be more likely to have a child with autism. But, he says, he chose to study Eindhoven after parents contacted him about an autism epidemic there, rather than, as some researchers may prefer, comparing the prevalence of autism in randomly selected IT regions with that of non-IT regions with similar demographics. And the Eindhoven school records did not disclose parental age or level of education — both of which are positively correlated with autism diagnoses — or whether the parents worked in the IT industry.

Indeed, researchers say that several other factors could explain the seeming correlation between autism and science or engineering. A 2010 analysis of autism diagnoses in California¹¹ did not find that autism clustered preferentially around areas rich in IT industry. Instead, it found that clusters tended to occur in areas where parents were older and educated to a higher level than were parents in surrounding areas. "Virtually all of these clusters were also clusters of higher education," says lead author Irva Hertz-Picciotto, an epidemiologist at the University of California, Davis.

People who have progressed further in education tend to have children later in life, and at least some evidence suggests that older parents are at higher risk of having children with autism. Parents who are more educated are also more likely to be aware of the symptoms of autism and to seek a diagnosis, which can open the door to support and education services. One Silicon Valley school for children with learning disabilities costs US\$30,000 per student per year, but if a child has been diagnosed with autism, the school district may pick up the tab.

In response to criticisms of his Eindhoven study, Baron-Cohen says he plans to follow up by looking at the age, occupation and other details of the parents, and that he'd also like

Despite the criticisms of Baron-Cohen's experiments, most of his colleagues commend him for putting his theories forward, and many are open to the possibility that parts of them could prove correct. "He does try to address big questions that many of us would be too wimpy to take on," says Lord.

Constantino is testing related ideas. He has developed the 'social responsiveness scale' — a questionnaire to measure autistic-like traits in the general population. He found hints that parents with more autistic-like traits tend to partner with each other, and that when they do, their children have even more of those traits than their parents¹². Those children, however, are not more likely to be diagnosed with autism¹³. What is needed now, Constantino says, is a large study that determines whether having two parents with autistic-like traits is more common among people with autism than in the general population. "Those are the kind of data one needs," he says, "rather than to infer, from an epidemiological cluster in a place where people tend to be a little nerdier, that that's why you've got more autism there."

For now, the idea that technical brilliance requires a dash of autism seems to have taken root, at least in some tech and science hubs. It's a trend that, for Happé, provokes mixed feelings. "On the one hand, I'm glad that 'geek chic' has some kudos in our current society, because a lot of people with AS or ASD have a jolly tough and unpleasant life, and if people can recognize their talents a little more, I'm glad for that."

On the other hand, she says, "a large number of children with autism have significant intellectual disabilities and no speech. For their parents to be surrounded by people spotting all these famous people and saying they have autism, it must be absolutely infuriating." ■

Lizzie Buchen is a freelance writer based in San Francisco.

"I do think that when these geeks marry each other, that's bad news for the offspring."

Baron-Cohen says that when he reanalysed the data and controlled for education level, he found that fathers of children with autism were still more likely to be engineers, although the difference was smaller.

One of Baron-Cohen's most recent studies comes from the town of Eindhoven, a technology hub in the Netherlands. By examining school records, he found that children living in the town were 2–4 times more likely to be diagnosed with autism than were children living in two other Dutch towns of

to examine autism rates in other IT centres, such as Silicon Valley. He's putting together a large online survey (go.nature.com/umyv61) to gather detailed information about the general population — including age, education, occupation and hobbies — to explore whether these factors correlate with having a child with autism. He says that Hertz-Picciotto's study didn't support his hypothesis because it "was not designed to look at autism in IT-rich regions. What I'm doing is coming at it in a different way," he says.

1. Baron-Cohen, S. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* **30**, 865–872 (2006).
2. Baron-Cohen, S. & Wheelwright, S. *Br. J. Psychiatry* **175**, 484–490 (1999).
3. Baron-Cohen, S., Richler, J., Bisarya, D., Guranathan, N. & Wheelwright, S. *Phil. Trans. R. Soc. Lond. B* **358**, 361–374 (2003).
4. Baron-Cohen, S., Wheelwright, S., Burtenshaw, A. & Hobson, E. *Hum. Nature* **18**, 125–131 (2007).
5. Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J. & Clubley, E. *J. Autism Dev. Disord.* **31**, 5–17 (2001).
6. Baron-Cohen, S., Wheelwright, S., Spong, A., Schill, V. & Lawson, J. *J. Dev. Learn. Disord.* **5**, 47–78 (2001).
7. Pellicano, E. *et al. Proc. Natl Acad. Sci. USA* **108**, 421–426 (2011).
8. Baron-Cohen, S., Wheelwright, S., Stott, C., Bolton, P. & Goodyer, I. *Autism* **1**, 153–163 (1997).
9. Jarrold, C. & Routh, D. *A. Autism* **2**, 281–289 (1998).
10. Roelfsema, M. T. *et al. J. Autism Dev. Disord.* <http://dx.doi.org/10.1007/s10803-011-1302-1> (2011).
11. Van Meter, K. C. *et al. Autism Res.* **3**, 19–29 (2010).
12. Constantino, J. N. & Todd, R. D. *Biol. Psychiatry* **57**, 655–660 (2005).
13. Virkud, Y. V., Todd, R. D., Abbacchi, A. M., Zhang, Y. & Constantino, J. N. *Am. J. Med. Genet. B* **150B**, 328–334 (2009).



A voice for science

Convinced by the evidence that vaccines do not cause autism, Alison Singer started a research foundation that pledges to put science first.

The e-mail that ended one career for Alison Singer, but started another, arrived as she was cooking dinner for her daughters one evening in January 2009. Singer was preoccupied. At a committee meeting she was due to attend in Washington DC the next day, she and others were set to vote on a plan that would direct much of the United States' spending on autism research for the next year.

Singer, who had her laptop perched on the kitchen counter, immediately noticed the e-mail from another committee member — a mother who was convinced that vaccines had caused her son's autism. The message proposed last-minute language for inclusion in the plan, endorsing more research into

BY MEREDITH WADMAN

whether vaccines can trigger the disorder of communication and movement. Singer knew immediately that this would cause her serious difficulties. Having read the literature and talked to numerous scientists, she was convinced that no studies supported a link between autism and vaccines. But she was also the top communications executive at Autism Speaks in New York, autism's most prominent research and advocacy group. The organization supports vaccine-related research, and Singer knew that her bosses would expect her to vote for more studies of vaccines as a possible cause of the condition.

At 11:10 p.m., Singer hit 'send' on an e-mail of her own, to Bob and Suzanne Wright, the co-founders of Autism Speaks. "I've concluded

that as a matter of personal conscience, I cannot vote in favor of dedicating more funds to vaccine research that has already been undertaken and which I and many others find conclusive," her message read. "I feel compelled to offer my resignation."

With that, Singer became a solo operator in the world of autism-research funders. Within months, she would launch the Autism Science Foundation (ASF), a tiny New York-based charity with a relentless focus on rigorous science, a niche supporting the youngest researchers and a guiding principle that "vaccines save lives; they do not cause autism."

The ASF is scarcely a blip on the big screen of autism-research spending: in this, its second full year of operations, it is awarding US\$220,000 in grants to young researchers; last year, it spent \$180,000. The two major

ANTONIO BOLFO/GA BY GETTY IMAGES

non-governmental players in US autism research, the Simons Foundation in New York and Autism Speaks, last year spent \$54 million and \$21 million, respectively (see 'All change for autism').

But Singer's charity is drawing notice as much for the aims and quality of its work as for its magnitude. In August, GuideStar, a major charity-rating group based in Washington DC, singled out the ASF as a "promising start-up", calling it "a shining star to those interested in real science and evidence based interventions".

The fledgling foundation has also won endorsements from leaders at the American Academy of Pediatrics in Elk Grove Village, Illinois; the National Institutes of Health (NIH) in Bethesda, Maryland; and the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia. "The Autism Science Foundation is an important voice for scientific direction in the autism community," says Coleen Boyle, the director of the National Center on Birth Defects and Developmental Disabilities at the CDC.

The jury is still out on whether Singer and her tiny organization can do much to combat the perception of a link between autism and vaccines that has become cemented in many minds. But if anyone is going to do it, Singer is a strong candidate. She "is a force of nature", says Thomas Insel, director of the National Institute of Mental Health in Bethesda, which spends more than \$80 million on autism each year — more than half of the NIH's autism budget. "I have enormous respect for her abilities."

"She is one of the most strong-willed people that I have ever met," adds Christie Buchovecky, an ASF-funded graduate student at Baylor College of Medicine in Houston, Texas, who has a 12-year-old cousin with autism. "And she somehow manages to do that while still being extremely caring and supportive. She knows where the parents are coming from."

ALL IN THE FAMILY

On Saturdays 40 years ago, when Singer (then Alison Tepper) was 5 years old and wanted to be in ballet class, she and her parents would set out on a very different errand: visiting her autistic 7-year-old brother, Steven, at the Willowbrook State School, an institution on Staten Island, New York, that housed more than 5,000 people diagnosed with mental retardation and developmental disabilities. Singer's parents had placed Steven in Willowbrook out of fear for their children's safety. Steven could be self-injurious and a danger to others; once, he had brought a whole wall unit,

including a television, crashing to the floor.

Willowbrook was a Dickensian institution that would eventually come to symbolize everything that was wrong with the way America cared for its mentally ill. In 1965, New York Senator Robert Kennedy alluded to the facility as "a snake pit". In 1972, an undercover exposé by local television channel WABC-TV showed residents sitting and staring vacantly; rocking, often naked, in overcrowded, barren rooms; or abandoned on bathroom floors. When Singer

"People didn't talk about it then. There was a tremendous stigma associated with autism."

and her family visited Steven, they would see him only in the visitors' room. Still, the stench of urine bothered Singer. So did the noise. "There was a lot of screaming," she recalls. "I didn't like it."

In 1971, shortly before the documentary aired, the Tepper family moved from the New York borough of Queens to the suburbs, and transferred Steven to a nearby facility called Letchworth Village. In time, the family — Singer also has a younger brother — began visiting Steven less often. Singer's mother, who had been told she was to blame for Steven's condition, instructed her daughter to tell people that she had just one, younger, brother. "People didn't talk about it then. No one talked about it," Singer says. "There was a tremendous stigma associated with autism."

Singer carried her secret out of childhood as she set out to become a journalist. In 1993, having graduated with a degree in economics from Yale University in New Haven, Connecticut, and a master's from Harvard Business School in Boston, Massachusetts, she was hired as vice-president of programming for the cable division of television network NBC. Singer was charged with bringing news content to desktop computers through the new phenomenon called the Internet. She married in 1994; in 1997, her first daughter, Jodie, was born. Singer thought something was wrong, she says, "from the day she was born". Jodie cried constantly. She didn't want to eat. She wouldn't sleep. One diagnosis followed another: failure to thrive; early colic; late colic. "I just thought I didn't have the mommy gene," Singer recalls.

She started questioning her mother over and over. "Did Steven ever do this? Did Steven ever behave like this?" My mom would say: 'No, Steven never did that.' Singer also took heart from a conversation with an obstetrician who assured her that autism is "not genetic". Still, as Jodie became a toddler, Singer remained deeply

troubled by her inability to communicate. "She used to recite the *Madeline* book from beginning to end perfectly," says Singer. "But if I said, 'Do you want juice?' she couldn't answer yes or no."

Singer called the state department of health several times to schedule an assessment for Jodie. Each time, in a rush of ambivalence, she would call back and cancel. Finally, she made a firm appointment. When a psychologist and a case worker visited her at home, Jodie was two

years, eight months old. (Singer had just given birth to Jodie's healthy younger sister, Lauren.) Singer expected them to say that nothing was wrong. Instead, they diagnosed Jodie as severely autistic, telling Singer gravely, "Don't worry, we're going to get her help." When they were gone, says Singer, "I just bawled".

Then, she says, "I pulled myself together".

DRIVING ON

'Together' seems an apt word for Singer. She comes across as organized, articulate and, above all, driven — although not without a sense of humour. On a September day, Singer showed this reporter around what she jokingly referred to as "the worldwide headquarters" of the ASF: a cubicle on the fourth floor of a lower Manhattan office building. There, Singer's one paid member of staff — Jonathan Carter, a recent college graduate with an autistic brother — was labouring over the just-relaunched ASF website. Singer herself does not draw a salary.

It's a long way from the high-flying position she held at Autism Speaks, where in her last year she made \$187,000. Bob and Suzanne Wright had made Singer their first employee when they launched the group in 2005, shortly after their grandson was diagnosed with autism. They were well placed to make things happen: Bob Wright was chief executive of media conglomerate NBC Universal and vice-chairman of General Electric, NBC's parent company. During the group's first months, when Singer served as interim chief executive, she, Suzanne Wright and an assistant constituted the entire staff. Working with a shared passion and purpose, from an office on the 51st floor of NBC Universal's headquarters, Singer and Suzanne Wright became fast friends. "Suzanne doesn't take no for an answer and neither do I. So we got a lot done," Singer recalls.

Their completed to-do list included landing coverage for Autism Speaks on television



THE AUTISM ENIGMA

Sorting fact from fiction
nature.com/autism

programmes *Oprah* and *Larry King Live*, and in the pages of *The Wall Street Journal*. Singer turned the organization's website into the most highly trafficked in the autism world. The group, she later wrote in her resignation e-mail, "elevated 'autism' to the global vocabulary". In 2008, Singer's final year with Autism Speaks, the charity raised more than \$70 million.

But there was friction, too. The Wright's daughter, Katie, fervently believed that vaccines were the cause of her son's autism, and was pushing her parents to fund more vaccine research. This didn't sit well with Singer. Although she had been open to such research when Jodie was first diagnosed, Singer had followed the growing number of studies that debunked any link between autism and the once-suspect measles, mumps and rubella vaccine, or between autism and thimerosal, a mercury-containing preservative that was used in some early-childhood vaccines until 2001. For the past several years, she says, the data have been very clear. "There were no studies indicating a link between autism and vaccines." As time went on, Singer became increasingly uncomfortable with her organization's continued funding of vaccine-related research. She was even more disturbed by its failure to state, publicly and unequivocally, that all the data show that vaccines do not cause autism.

The committee meeting in January 2009 brought matters to a head. The Interagency Autism Coordinating Committee, a group of federal-agency and public representatives, was charged with developing an annual strategic plan for autism research, which would be used to guide work done by federal agencies, both internally and in collaboration with private foundations. Singer refused to support a plan that would broadly sanction vaccine studies. "She knew that her resignation would be seen as a desertion," says Paul Offit, chief of infectious diseases at the Children's Hospital of Philadelphia in Pennsylvania, and now a close colleague of Singer's and a member of the ASF's board of directors. "It was a very, very hard thing for her to do."

The day after she resigned, Singer spoke at the committee meeting, letting members know how she was going to vote, and why. There were two votes related to vaccine research at the meeting. The first, on whether to add the language that had circulated the previous evening, failed. The committee also voted on whether to cut two existing vaccine-research objectives from the plan. All the other members representing advocacy groups or affected families opposed removing the objectives; Singer and the federal-agency scientists did not, and the objectives remained. Autism Speaks immediately put out a press release distancing itself from Singer and withdrawing its support for federal autism-research plan.

Suzanne Wright has not contacted Singer, or replied to e-mails from her, since Singer resigned. But Singer did receive many supportive messages. "I understand your reasons completely and agree with you wholeheartedly," wrote one supporter, Mark Krinsky, whose son, now 25, has autism. "What are your plans? Any group would be lucky to get you but maybe you should start a new one. I will volunteer."

It was letters and phone calls such as these, says Singer, "that got me thinking about starting a new advocacy group". She launched the ASF in March 2009.

PERSONAL CHALLENGE

Singer, who drives a blue Honda minivan decorated with bumper stickers reading "Autism Science Foundation" and "Vaccinate your baby", spends two days a week working from her home in Scarsdale, an upmarket northern suburb of New York. This allows her to juggle a mother's duties with her work. One Monday afternoon, she arrives home from the office to find Lauren, now 11, downstairs finishing her homework, and Jodie, 14, upstairs with Keith Amerson, one of her therapists. Amerson, who is an expert in a leading autism therapy called applied behaviour analysis (ABA), has been working with Jodie since she was diagnosed.

Today, Jodie is sitting beside Amerson at a table, in front of two plastic plates. The yellow one has one piece of toy food on it; the green one has several. "Jodie, which one has more?" asks Amerson. Jodie points to the yellow plate.

When her mother comes in, Jodie lights up. The two bounce on the bed in Jodie's bedroom

and sing a song, that, Singer says, they have sung "a thousand times". Jodie: "Say: 'I love my Jodie.'" Singer: "I love my Jodie."

But Jodie is challenging. She sleeps for as little as four hours a night. Every door and window in the house has had an alarm ever since, in the wee hours one winter morning, shoeless and clad in pyjamas, she set off down the driveway in search of an egg roll at a Chinese restaurant. Luckily, Singer was awake and stopped her.

When Amerson pushes Jodie to find the prices beside pictures of root beer and cake on a sheet of paper, and enter them in a calculator, Jodie cries "Oh no!" and runs her flapping, nervous hands through her brown pony tail. "Do you need some help?" asks Amerson. "Keith!" replies Jodie. "Say: 'I need some help,'" counters Amerson. "I need some help," Jodie says.

The goal of ABA therapy, says Amerson, "is to make her as independent as possible and to improve the quality of her life". A task as quotidian as writing '\$1.74' legibly, with the decimal point in the right place, is a large challenge, its achievement a victory. "It's very tough work," he says.

YOUNG BRAINS

On 8 February 2010, almost a year after it launched, the ASF awarded its first grants, splitting \$180,000 between six PhD students. Singer has chosen to focus on the youngest scientists because, she says, "it fuels the pipeline for the future. You are encouraging that person to pursue autism research as opposed to research in some other disease." The foundation boosted its funding by more than 20% this year, handing



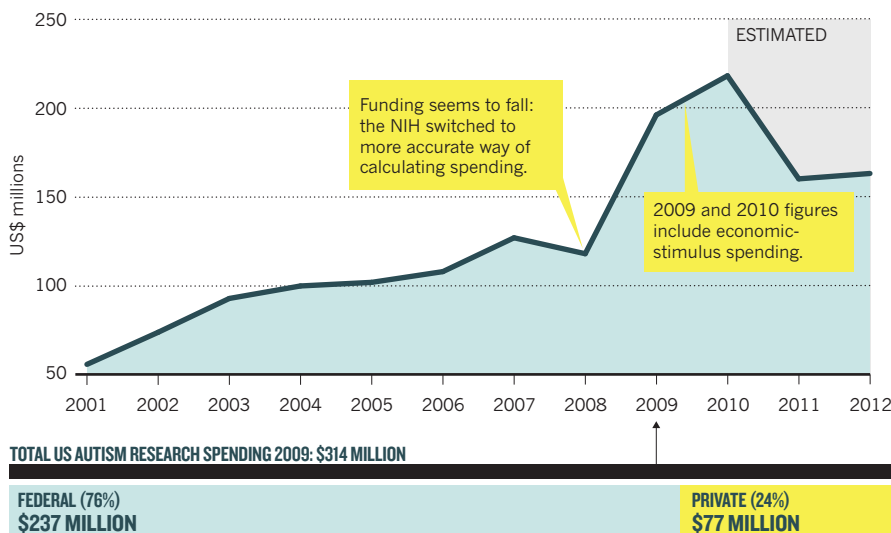
Alison Singer (right) in 1968. Her brother Steven (left) has autism, as does her daughter Jodie.

A. SINGER

SOURCE: NIH

All change for autism

Funding for autism research from the US National Institutes of Health (NIH) has risen in recent years (top). In 2009, federal funding dwarfed that from private sources (bottom), although this is also rising.



out a further six awards to graduate students and postdoctoral researchers.

Chosen by a ten-member scientific advisory board, the awards run the gamut from basic science to diagnosis and treatment. One 2011 recipient, Jill Locke at the University of Pennsylvania and the Children's Hospital of Philadelphia, is working to translate behavioural techniques proven in the research setting to Philadelphia's public schools. She is training school personnel to improve the social interactions between children with autism and their playground peers. Another recipient is Haley Speed, a postdoctoral fellow at the University of Texas Southwestern Medical Center in Dallas, who is building on work that she recently co-authored in *Cell*, describing a mouse with a mutation in the *Shank3* gene, which is linked to autism in humans (M. A. Bangash *et al.* *Cell* **145**, 758–772; 2011). In the past, Speed had failed to win NIH postdoctoral awards, and she believes that the riskiness of her current research — in which she hopes to find compounds that could correct the animals' abnormal neuron-to-neuron communication and restore normal behaviour — would not help her case with conservative study sections at the biomedical agency. "The ASF had faith in us, whereas the NIH did not," she says.

Speed adds that it inspires her to know that \$10 and \$20 donations from affected families are paying for her work. "On a bad day, if your equipment breaks or your experiment fails — which they all do — it gives you an extra boost as you to go pick yourself up and do it all again."

Indeed, many donations to the ASF come in small cheques from affected families and

their friends, or larger ones from the proceeds of football goal-a-thons and charity motorbike rides. (A significant amount has also come from large donors, including Offit, who has donated all the royalties from the sales of his books challenging the anti-vaccine movement to ASF — an amount he estimates in tens of thousands of dollars.) Singer does not think that her opposition to vaccine research limits the group's ability to raise funds — far from it. She says the organization has mobilized the thousands of families who were sick of the autism story being hijacked by the vaccine hypothesis. "ASF is their voice."

The ASF is still dwarfed by the size and reach of other organizations supporting autism research. The Simons Foundation does not fund vaccine research, but Autism Speaks spends about 2% of its budget — nearly twice as much as the ASF spent on all research this year — on studies that are relevant to vaccines, says Geri Dawson, chief science officer for Autism Speaks and a professor of psychiatry at the University of North Carolina at Chapel Hill. "We are not funding any studies that directly address whether vaccines cause autism," says Dawson. "The evidence strongly suggests that there is not a link between autism and vaccines. What we are trying to understand through our research funding is the role of the immune system in autism, which certainly could be relevant to the question of vaccines."

For instance, Dawson says, Autism Speaks is funding studies investigating the idea that disorders in the cellular powerhouses, mitochondria, might influence responses to immune challenges such as infection and immunization. "We are willing to leave the

door open for the possibility of rare cases in which an immunization may have triggered the onset of autism symptoms due to an underlying medical or genetic condition."

Singer calls Dawson's words "a carefully crafted 'big tent' political statement designed to appeal to both sides of the issue" and an attempt "not to offend any potential donors on either side". Her concern, she says, is that statements suggesting that the jury is still out on autism and vaccines put children "at risk for vaccine-preventable disease".

In an e-mail, Katie Wright wrote that she "totally respects" Singer's beliefs about Jodie. "However, Alison doesn't live my life, doesn't know what caused my son's autism and wasn't with me the night he had a febrile seizure or the day he stopped talking. It is important not to tell other parents you know better than they do about their child's autism. We all love our kids very much, Alison and I certainly agree on that."

THE LONG HAUL

Although the world of autism research is expanding, horizons for many of those with the condition have scarcely changed. Singer's brother, Steven, lives in a group home in Rockland County, not far from Scarsdale. Now 47, he attends a day programme and helps to deliver Meals on Wheels three mornings a week. Still, he has never spoken, and needs round-the-clock supervision. "If he had had early intervention when he was two, when his brain was more malleable, then who knows where he would be today," says Singer, who visits him several times a year. "But I'm a big believer that you never stop trying."

Singer is not deterred by the sometimes agonizingly slow pace of science, or by the stark fact that the genetics and aetiology research that her foundation is funding is unlikely to ever directly benefit Steven or Jodie. Singer now believes that her daughter's autism was largely caused by genes, but genetic testing when she was first diagnosed revealed no known pathogenic deletions or duplications in her genome. Singer had Jodie retested earlier this year, after a glut of studies identified mutations implicated in autism. The tests still came up empty.

Nonetheless, Singer believes that the ASF's research into treatment will improve the lives of people who have autism today — including Jodie. She also focuses on her other daughter, Lauren, who has already asked whether she, too, will have a child with autism.

Says Singer: "I would like to be able to answer her with: 'If you do, we will know how to help.'" ■ SEE EDITORIAL P.5

Meredith Wadman is a reporter for *Nature* in Washington DC.

COMMENT

STATISTICS Global standards are needed for weighing evidence in court **p.36**

LITERATURE Lewis Carroll's experiments with time in the Alice books **p.38**

PHYSICS The outsiders who imagine alternative ideas of the Universe **p.40**



OBITUARY Steve Jobs and his aesthetic vision for the computing industry **p.42**

JESSICA DIMMOCK/VII



Working with autistic scientist Michelle Dawson (right) has helped Laurent Mottron (left) change his entire perception of the condition.

The power of autism

Recent data — and personal experience — suggest that autism can be an advantage in some spheres, including science, says **Laurent Mottron**.

Most grant proposals, research papers and reviews on autism open with, “Autism is a devastating disorder”. Mine do not.

I am a researcher, clinician and lab director concentrating on the cognitive neuroscience of autism. Eight autistic people have been associated with my group: four research assistants, three students and one researcher.

Their roles have not been limited to sharing their life experiences or performing mindless data entry. They are there because of their intellectual and

personal qualities. I believe that they contribute to science because of their autism, not in spite of it.

Everyone knows stories of autistics with extraordinary savant abilities, such as Stephen Wiltshire, who can draw exquisitely detailed urban landscapes from memory after a helicopter tour. None of my lab members is a savant. They are ‘ordinary’



THE AUTISM ENIGMA
Sorting fact from fiction
nature.com/autism

autistics, who as a group, on average, often outperform non-autistics in a range of tasks, including measures of intelligence.

As a clinician, I also know all too well that autism is a disability that can make daily activities difficult. One out of ten autistics cannot speak, nine out of ten have no regular job and four out of five autistic adults are still dependent on their parents. Most face the harsh consequences of living in a world that has not been constructed around their priorities and interests.

But in my experience, autism can also be an advantage. In certain settings, ►

► autistic individuals can fare extremely well. One such setting is scientific research. For the past seven years, I have been a close collaborator of an autistic woman, Michelle Dawson. She has shown me that autism, when combined with extreme intelligence and an interest in science, can be an incredible boon to a research lab.

I first met Dawson when we were interviewed together for a television documentary about autism. Some time later, after disclosing to her employers that she was autistic, she experienced problems in her job as a postal worker and so had learned everything about how the legal system deals with employees with disabilities. I recognized her skill for learning and asked her to become a research assistant in my lab. When she edited some of my papers, she gave exceptional feedback and it was clear that she had read the entire bibliography. The more she read, the more she learned about the field. Almost ten years ago, I offered her an affiliation to the lab. We've now co-authored 13 papers and several book chapters.

TESTING ASSUMPTIONS

Since joining the lab, Dawson has helped the research team question many of our assumptions about and approaches to autism — including the perception that it is always a problem to be solved. Autism is defined by a suite of negative characteristics, such as language impairment, reduced interpersonal relationships, repetitive behaviours and restricted interests. Autism's many advantages are not part of the diagnostic criteria. Most educational programmes for autistic toddlers aim to suppress autistic behaviours, and to make children follow a typical developmental trajectory. None is grounded in the unique ways autistics learn.

In cases where autistic manifestations are harmful — when children bang their heads on the walls for hours, for example — it is unquestionably appropriate to intervene. But often, autistic behaviours, although atypical, are still adaptive.

For instance, one sign of autism is using another person's hand to ask for something,

such as when a child places her mother's hand on the refrigerator to ask for food, or on the door knob to ask to go outside. This behaviour is unusual, but it lets children communicate without language.

Even researchers who study autism can display a negative bias against people with the condition. For instance, researchers performing functional magnetic resonance imaging (fMRI) scans systematically report changes in the activation of some brain regions as deficits in the autistic group — rather than evidence simply of their alternative, yet sometimes successful, brain organization.

Similarly, variations in cortical volume have been ascribed to a deficit when they appear in autism, regardless of whether the cortex is thicker or thinner than expected¹.

When autistics outperform others in certain tasks, their strengths are frequently viewed as compensatory of other deficits, even when no such deficit has been demonstrated empirically.

Without question, autistic brains operate differently. Most notably, they rely less on their verbal centres. When non-autistic people look at an image of a saw, for example, their brains are activated in regions that process both visual information and language.

In autistics, there is comparatively more activity in the visual-processing network than in the speech-processing one², and this seems to be a robust characteristic of autism, across a wide array of tasks³. This redistribution of brain function may nonetheless be associated with superior performance⁴ (see fMRI images below).

These differences may also have downsides, such as difficulties with spoken language. But they can confer some advantages. A growing body of research is showing that autistics outperform neurologically

“Too often, employers don't realize what autistics are capable of, and assign them repetitive, almost menial tasks.”

typical children and adults in a wide range of perception tasks, such as spotting a pattern in a distracting environment⁵.

Other studies have shown that most autistic people outperform other individuals in auditory tasks (such as discriminating sound pitches⁶), detecting visual structures⁷ and mentally manipulating complex three-dimensional shapes. They also do better in Raven's Matrices, a classic intelligence test in which subjects use analytical skills to complete an ongoing visual pattern. In one of my group's experiments, autistics completed this test an average of 40% faster than did non-autistics⁴.

A CHANGED MIND

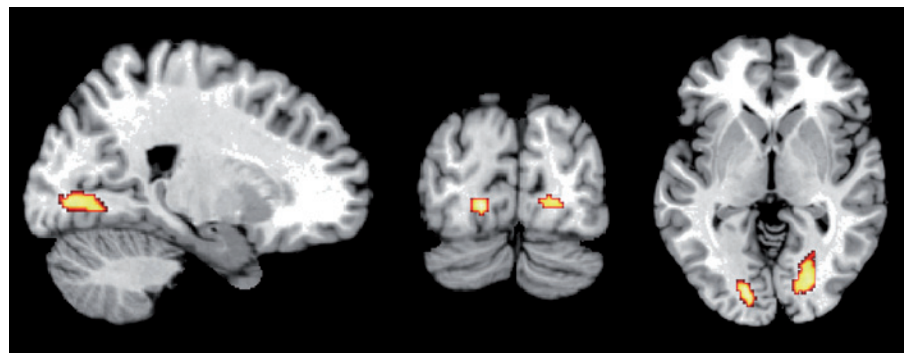
A few years ago, my colleagues and I decided to compare how well autistic and non-autistic adults and children performed in two different types of intelligence test: non-verbal ones, such as Raven's Matrices, that need no verbal instructions to complete, and tests that rely on verbal instructions and answers. We found that non-autistics as a group performed consistently in both types of test — if they scored in the 50th percentile in one, they tended to score around the 50th percentile in the other. However, autistics tended to score much higher in the non-verbal test than in the verbal one (see 'Autistic intelligence') — in some cases, as many as 90 percentile points higher⁸.

Despite autistics' success in Raven's Matrices, I, too, used to believe that verbal tests were the best measures of intelligence. It was Dawson who opened my eyes to this 'normocentric' attitude. She asked me: if autistics excel in a task that is used to measure intelligence in non-autistics, why is this not considered a sign of intelligence in autistics?

It is now amazing to me that scientists continue to use, as they have for decades, inappropriate tests to evaluate intellectual disability among autistics, which is routinely estimated to be about 75%. Only 10% of autistics have an accompanying neurological disease that affects intelligence, such as fragile-X syndrome, which renders them more likely to have an intellectual disability.

I no longer believe that intellectual disability is intrinsic to autism. To estimate the true rate, scientists should use only those tests that require no verbal explanation. In measuring the intelligence of a person with a hearing impairment, we wouldn't hesitate to eliminate components of the test that can't be explained using sign language; why shouldn't we do the same for autistics?

Of course, autism affects other functions, such as communication, social behaviour and motor abilities. These differences can render autistics more dependent on others, and make everyday life much more difficult. None of my arguments above is intended to minimize that.



For certain tasks, autistics use their brains differently: these fMRI images depict the perceptual regions of the brain activated more among autistics than non-autistics during a non-verbal intelligence test.

SOURCE: REF. 4

Too often, employers don't realize what autistics are capable of, and assign them repetitive, almost menial tasks. But I believe that most are willing and capable of making sophisticated contributions to society, if they have the right environment. Sometimes the hardest part is finding the right job — but organizations are now arising to address this problem. For example, Aspiritech, a non-profit organization based in Highland Park, Illinois, places people who have autism (mainly Asperger's syndrome) in jobs testing software (www.aspiritech.org). The Danish company Specialisterne has helped more than 170 autistics obtain jobs since 2004. Its parent company, the Specialist People Foundation, aims to connect one million autistic people with meaningful work (www.specialistpeople.com).

Many autistics, I believe, are suited for academic science. From a young age, they may be interested in information and structures, such as numbers, letters, mechanisms and geometrical patterns — the basis of scientific thinking⁹. Their intense focus can lead them to become self-taught experts in scientific topics. Dawson, for example, does not have a scientific degree, but she has learned and produced enough in a few years of reading neuroscience papers to conduct certain types of research. At this point, she deserves a PhD.

INSTANT RECALL

Research has consistently shown that, on average, autistics present strengths that can be directly useful in research. They can simultaneously process large pieces of perceptual information, such as large data sets, better than non-autistics can¹⁰. They often have exceptional memories: most non-autistic people can't remember what they read ten days ago; for some autistics, that's an effortless task. Autistic people are also less likely to misremember data. This comes in handy in science: whereas the methodologies used in studies of face-perception in autism are for me terribly similar, Dawson can instantaneously recall them.

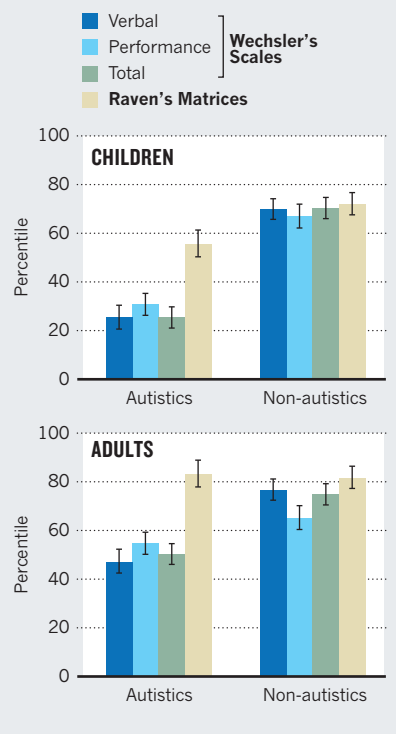
Many autistics are good at spotting recurrent patterns in large amounts of data, and instances in which those patterns have been broken. In my lab, Dawson noticed a discrepancy in the standards applied to various types of treatments: to develop a drug, researchers must conduct elaborate studies including randomized controlled trials, but this is not a requirement for behavioural interventions for autistics, despite the huge costs of such interventions (up to US\$60,000 a year for each individual) and their potential negative consequences.

It is thus worrying that some countries,

NATURE.COM
Read Dawson's
essay on behavioural
therapies:
go.nature.com/c6d9mj

AUTISTIC INTELLIGENCE

Non-autistics typically perform equally well in tests of verbal and non-verbal intelligence. Autistics, however, score much higher in non-verbal tests, such as Raven's Matrices, than in verbal ones, such as Wechsler's Scales.



including France, have proposed mandatory interventions that aim to get autistics to adopt 'typical' learning and social behaviours, which have not been tested using the standards applied to other areas in science.

Dawson's keen viewpoint also keeps the lab focused on the most important aspect of science: data. She has a bottom-up heuristic, in which ideas come from the available facts, and from them only. As a result, her models never over-reach, and are almost infallibly accurate, but she does need a very large amount of data to draw conclusions. By contrast, I have a top-down approach: I grasp and manipulate general ideas from fewer sources, and, after expressing them in a model, go back to facts supporting or falsifying this model. Combining the two types of brains in the same research group is amazingly productive.

Because data and facts are paramount to autistic people, they tend not to get bogged down by the career politics that can sidetrack even the best scientists. They prefer not to seek popularity, promotions or vast numbers of papers; they may post their best ideas on the web rather than publish them.

In 2004, Dawson gained recognition within the autistic community and among autism researchers and clinicians after posting online an essay detailing the ethical shortcomings of the intensive behavioural

therapies used with autistic children.

Of course, autistics will not thrive in all careers. Given their social differences, they will often struggle in people-oriented fields, such as retail or customer service. Ideally, autistic individuals would have mediators who could help settle situations that trigger anxiety in them — typically anything unscheduled or hostile, such as changes to an existing plan, computer problems or negative criticism.

Despite these caveats, Dawson and other autistic individuals have convinced me that, in many instances, people with autism need opportunities and support more than they need treatment. As a result, my research group and others believe that autism should be described and investigated as a variant within the human species. These variations in gene sequence or expression may have adaptive or maladaptive consequences, but they cannot be reduced to an error of nature that should be corrected.

The hallmark of an enlightened society is its inclusion of non-dominant behaviours and phenotypes, such as homosexuality, ethnic differences and disabilities. Governments have spent time and money to accommodate people with visual and hearing impairments, helping them to navigate public places and find employment, for instance — we should take the same steps for autistics.

Scientists, too, should do more than simply study autistic deficits. By emphasizing the abilities and strengths of people with autism, deciphering how autistics learn and succeed in natural settings, and avoiding language that frames autism as a defect to be corrected, they can help shape the entire discussion. ■

Laurent Mottron is professor of psychiatry and holds the Marcel & Rolande Gosselin Research Chair in Cognitive Neuroscience of Autism at the University of Montreal. He is also director of the autism programme at the Hospital Rivière-des-Prairies, 7070 boul. Perras, Montreal H1E 1A4, Quebec, Canada.
e-mail: mottronl@uniserve.com

- Gernsbacher, M. A. *Observer* **20**, 43–45 (2007).
- Gaffrey, M. S. et al. *Neuropsychologia* **45**, 1672–1684 (2007).
- Samson, F., Mottron, L., Soulières, I. & Zeffiro, T. A. *Hum. Brain Mapp.* <http://dx.doi.org/10.1002/hbm.21307> (2011).
- Soulières, I. et al. *Hum. Brain Mapp.* **30**, 4082–4107 (2009).
- Pellicano, E., Maybery, M., Durkin, K. & Maley, A. *Dev. Psychopathol.* **18**, 77–98 (2006).
- Heaton, P. J. *Child Psychol. Psych.* **44**, 543–551 (2003).
- Perreault, A., Gurnsey, R., Dawson, M., Mottron, L. & Bertone, A. *PLoS ONE* **6**, e19519 (2011).
- Dawson, M., Soulières, I., Gernsbacher, M. A. & Mottron, L. *Psychol. Sci.* **18**, 657–662 (2007).
- Mottron, L., Dawson, M. & Soulières, I. *Phil. Trans. R. Soc. Lond. B* **364**, 1385–1391 (2009).
- Plaisted, K., O'Riordan, M. & Baron-Cohen, S. *J. Child Psychol. Psych.* **39**, 765–775 (1998).



Sally Clark was cleared of murder in 2003 after statistical evidence was found to be flawed.

Improve statistics in court

Experts must agree a set of acceptable ways to assess and present forensic evidence, says **Norman Fenton**.

Misuse of statistics is a blight on the law. Verdicts have been influenced by incorrect statistical reasoning in dozens of widely documented cases^{1,2}. Sally Clark, for instance, was wrongly convicted of the murder of her two infant sons in a British court in 1999. She was cleared in 2003 after further investigation revealed that the probability of cot death had been calculated wrongly as being far too low. Many more cases go unnoticed.

Most common fallacies of statistical reasoning can be avoided by applying Bayes' theorem, a rule that allows the evidence to be weighted. Yet the Bayesian approach is widely misunderstood and mistrusted in court. A year ago a UK Court of Appeal ruling (known as *Regina v. T.*)³, dealt a further blow. Quashing a murder conviction in which the prosecution had relied heavily on Bayesian methods to present footwear-matching evidence, the judge said that Bayesian methods were an inadmissible way to present expert evidence — except for DNA and “possibly other areas where there is a firm statistical base”. Such restrictions are a backward step for justice. The consequences will be that expert evidence will be

misinterpreted and widely suppressed.

Forensic, statistical and legal experts around the world have reacted to the *Regina v. T.* ruling with concern and criticism⁴. To confront such challenges, I am setting up an international consortium of statisticians, forensic scientists and academic and practising lawyers (80 people signed up in the first 2 months) to develop guidelines for when and how Bayesian reasoning should be used to present evidence. We will agree initial objectives in a workshop in London in December (see go.nature.com/agn3or).

UPDATED ODDS

Bayes' theorem is the accepted rule for updating the probability of a hypothesis given new evidence. Importantly, the formula can be used to weight the impact of pieces of evidence individually or in combination. Suppose, for example, that blood found at the scene of a crime is of a type that is prevalent in one in every thousand people, and the defendant has the same blood type. Clearly, the match increases the probability that it was the defendant's blood at the scene. But by how much?

To answer this we need to compare the prosecution likelihood (the probability of seeing the evidence if the prosecution hypothesis is true) with the defence likelihood (the probability of seeing the evidence if the defence hypothesis is true). In this example, the former would be equal to one, and the latter would be one in a thousand. So we are 1,000 times more likely to observe the evidence if the prosecution hypothesis is true than if the defence hypothesis is true.

A simple measure of the impact of evidence is the likelihood ratio, the prosecution likelihood divided by the defence likelihood (1,000 in this case). Values above one favour the prosecution (the higher the better); those below one favour the defence (the lower the better). A value of exactly one means that the evidence is worthless (the prosecution and defence are affected equally). The likelihood ratio is extremely valuable, but to draw definitive conclusions we need Bayes' theorem to tell us how the odds change when new evidence is added: our updated (posterior) odds equal the prior odds for the prosecution hypothesis multiplied by the likelihood ratio.

So if, for example, there were 10,000 other adults who could have been at the scene of the crime, our prior odds would have been 10,000 to 1 against the prosecution hypothesis. Once we see the evidence, the revised odds still favour the defence, but they have dropped to 10 to 1 against the prosecution (equivalently the probability that the defendant was not at the scene has gone from 99.99% to about 91%).

A common error — known as the ‘prosecutor fallacy’ — is to assume that the (revised) probability of the defence hypothesis is the same as the defence likelihood. A prosecutor might state, for example, that ‘the probability that the defendant was not at the scene given this evidence is one in a thousand’, when actually it is 91%. This is one of the most common statistical legal mistakes.

SUMMARY

- Statistical methods for weighing evidence are being blocked in recent court rulings
- Lawyers may be unwilling to quantify subjective evidence, preventing legal conclusions from being drawn
- The difficulty of presenting complex probability calculations hinders their widespread acceptance
- An international consortium of statisticians, lawyers and forensic scientists is drawing up guidelines for the use of statistics in court

Such an error might be spotted through intuitive reasoning that tallies with the Bayes result: of the 10,000 other adults, about 10 should have the same blood type as the defendant, so the blood match tells us that the defendant is one of 11 who could have been at the scene (see 'Bayesian reasoning'). However, the explanation is rarely so simple.

LEGAL RESISTANCE

Despite its potential utility, Bayes' theorem is not trusted by much of the legal profession. There are two main reasons: subjectivity is often misunderstood, and Bayesian arguments are difficult to present in a way that everyone will comprehend.

Subjective judgement about uncertain evidence is at the heart of the jury trial process. And most lawyers are fiercely resistant to the idea that a numerical figure can be attached to this uncertainty. Suppose, for example, that a defendant was known to be part of a mob, one of whom committed an attack. An eyewitness for the prosecution estimates there were 50 other people in the mob. So the odds that the defendant committed the attack are 50 to 1 against. But this is subjective — another witness for the defence might claim there were 100 other people in the mob.

Bayesians would seek to assess such subjective information by considering a range of odds, from 50 to 1 to 100 to 1, which is acceptable to both sides. Few lawyers would use such subjective numbers. But that often means that they cannot make important conclusions. For example, if further evidence arose, say with a likelihood ratio (as in the blood-match example) of 1,000, then the revised odds would range from 20 to 1 to 10 to 1, and strongly favour the prosecution. Similarly, in a medical negligence case in 2010, I showed that the claimant's argument was favoured irrespective of the evidence supplied by both sides⁵.

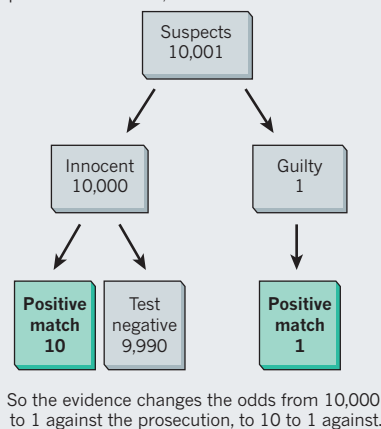
Misunderstandings of subjectivity have also restricted the types of evidence to which lawyers assume that Bayes and likelihood ratios can be applied. The requirement of a 'firm statistical base' as ruled by the judge in *Regina v. T.*, for example, has been interpreted in many cases as meaning that no subjective data can be used. But all probabilities — including DNA matches from the most comprehensive database — involve some subjective judgement⁶. It is better to acknowledge that subjectivity is inevitable and evaluate it using Bayes' theorem than to forego the evidence.

The fact that only the simplest Bayesian arguments can be explained from first principles in a way that laypeople will comprehend also limits their acceptance in court. A

▶ NATURE.COM
See our News
special on science
in court:
go.nature.com/ez6pwk

BAYESIAN REASONING

Although only 1 of 10,001 suspects for a hypothetical crime is guilty, 11 would have a positive match for a blood type with a prevalence of 1 in 1,000.



real case may need to make complex assumptions or include multiple pieces of dependent evidence, and these cannot be relayed in straightforward decision trees. Most Bayesian calculations are so complicated that software tools are needed to complete them⁷. So, in the case of *Regina v. Adams*⁷, even when the defence expert presented the Bayesian calculations (balancing subjective evidence for the defence with the prosecution's DNA-match probability) to the jury from first principles, the exercise wasn't successful. It also backfired when the appeal judge ruled against such use of Bayes' theorem in future trials².

CONSENSUS NEEDED

The *Regina v. T.* ruling has drawn fierce criticism from many experts (including some lawyers) who appreciate the benefits of Bayes' theorem. They regard the ruling as a constraint on accepted scientific practice⁵; although it has also been praised as an attempt to rule out probabilistic forensic evidence that is based on 'unscientific' data. Lacking a definition of what is 'scientific', lawyers are erring on the side of caution and rejecting the use of likelihood ratios, even in areas in which their use used to be standard (such as in fibre and glass matching). Experts are then left to make vague assertions about how well the evidence supports a hypothesis.

There have been isolated attempts to improve the understanding of probability within the law. Last year, for instance, the UK Royal Statistical Society's statistics and law working group issued guidance aimed at judges, lawyers, forensic scientists and expert witnesses⁸. But legal practice will change only when a critical mass of international experts, supported by key members of the judiciary, reaches a consensus on two points: when Bayesian reasoning about evidence can and cannot be applied;

and how to get it accepted in court without having to present the calculations from first principles.

Wider acceptance of Bayesian analysis also requires lawyers, expert witnesses and others to understand that there is a crucial difference between the genuinely disputable (subjective) prior assumptions, and the (objective) Bayesian calculations required to compute the conclusions from the different disputed assumptions. Lay observers must accept that they can question only the assumptions that go into the Bayesian calculations and not the calculations themselves. By considering ranges of subjective assumptions we can address the most persistent objections to the use of the theorem. Acceptance of emerging Bayesian software tools will remove the need to go through the calculations in court from first principles.

Proper use of probabilistic reasoning has the potential to improve the efficiency, transparency and fairness of the criminal justice system. Bayesian reasoning can help experts to formulate accurate and informative opinions; courts to determine the admissibility of evidence and identify which cases should and should not be pursued; and lawyers to explain, and jurors to evaluate, the weight of evidence during a trial. It can also help to identify any errors and unjustified assumptions in expert opinions.

There is still widespread disagreement about the type of evidence to which Bayesian reasoning should be applied and how it should be presented. There are ways to overcome these technical barriers, but cultural barriers still remain between the fields of science and law, and these will be broken down only by achieving a critical mass of relevant experts and stakeholders, united in their objectives. The international consortium is building towards such a consensus. ■

Norman Fenton is professor of risk information management at Queen Mary, University of London and chief executive of Agena, Cambridge, UK.
e-mail: norman@eecs.qmul.ac.uk

1. Aitken, C. G. G. & Taroni, F. *Statistics and the Evaluation of Evidence* 2nd edn (John Wiley & Sons, 2004).
2. Fenton, N. E. & Neil, M. *Aust. J. Legal Phil.* (in the press); available at http://www.eecs.qmul.ac.uk/~norman/papers/fenton_neil_prob_fallacies_June2011web.pdf
3. *Regina v. T.* [2010] available at <http://www.bailii.org/ew/cases/EWCA/Crim/2010/2439.pdf>
4. Berger, C. E. H., Buckleton, J., Champod, C., Evett, I. & Jackson, G. *Sci. Justice* **51**, 43–49 (2011).
5. Fenton, N. & Neil, M. *J. Biomed. Inform.* **43**, 485–495 (2010).
6. Thompson, W. C., Taroni, F. & Aitken, C. G. G. *J. Foren. Sci.* **48**, 47–54 (2003).
7. *Regina v. Adams* [1996] 2 Cr App R 467, [1996] Crim LR 898, CA and *Regina v. Adams* [1998] 1 Cr App R 377
8. Aitken, C., Roberts, P. & Jackson, G. *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings* (Royal Statistical Society, 2010).



The White Rabbit and his watch spark the temporal theme in *Alice's Adventures in Wonderland*.

MATHEMATICS

Alice in time

Gillian Beer chronicles the passage of time in its many manifestations through Lewis Carroll's enduring classics.

Time haunts both Alice books. Lewis Carroll, author of *Alice's Adventures in Wonderland* (1865) and *Through the Looking-Glass* (1871), was also Charles Dodgson, mathematician and logician, and so was aware of the disturbing arguments, new in the mid-nineteenth century, that suggested our view of the geometry of space and time was not universal.

As Dodgson, he was a devout Euclidean, believing that planes are flat and parallel lines never meet. As Lewis Carroll, he stepped across those boundaries.

Carroll saw problems of temporality as fundamental both to logic and to

possible worlds. In *Alice's Adventures in Wonderland*, time sparks the whole adventure. When Alice sights the White Rabbit at the start, the animal mutters about lateness, but it is the timepiece that startles her: "When the Rabbit actually *took a watch out of its waistcoat-pocket* [Carroll's italics], and looked at it... Alice started to her feet".

Belatedness, anxiety and physical props such as the watch bespeak the individual under the cosh of time-regulated society. Watches had become established as a token of human respectability and, along with the factory clock, were the instruments that

controlled industrialized labour. Carroll was a railway enthusiast, and the Alice books appeared when railway timetables required the regularizing of time across Britain.

During Carroll's lifetime, space and time came to be understood more and more as linked concepts. Chronometers kept time at sea and helped in the mapping of colonial claims, bringing time and space together. The new technology of the photograph, of which Carroll was an early adept, froze and made portable a moment and a place. And, as Jimena Canales pointed out in her book *A Tenth of a Second* (University of Chicago Press; 2010), that particular time unit had new significance, especially in measuring the speed of thought and reaction time.

German physicist and physiologist Hermann von Helmholtz was in the vanguard of such discoveries — including the measurement of the speed of the nerve impulse — from the 1850s. In an 1870 *Academy* article, 'Axioms of Geometry', he summarized many non-Euclidean insights of the previous two decades, such as the possibility of parallel lines intersecting. He ended by citing German mathematician Georg Riemann's "somewhat startling conclusion, that the axioms of Euclid may be, perhaps, only approximately true". In this spirit, Helmholtz asserts the logical congruity of conceiving intelligent beings living on squashed planets, or in two or four dimensions.

Carroll did not follow this non-Euclidean thinking professionally, but let it loose in his fiction. So Alice shrinks and swells, is crushed into the space of the Rabbit's house or finds her head swaying on an elongated neck in the canopy of a tree. In this alternative space and time, her body's shape is not constant and its relation to its environment is approximate. The child's everyday and helpless experience of growing, and of being always the wrong size in a world designed by adults, is meshed with new mathematical speculations.

But echoes and reflections of industrial, scientific and technological changes are not the only markers of temporality in these books: sundials, solar time and dreams each add their diverse processes.

The wayward non-causal sequences experienced in dreams nudge the episodes onward in both books. Dreams share with narrative the property of presenting experience as simultaneously in the past and yet in process. The games in the books — cards, croquet and chess — do not unroll within a rigid time frame. But they involve strategic moves, giving them a time-driven urgency. So in the looking-glass world, Alice does eventually become a queen when, as a pawn, she reaches the end of the chess board. But that purposeful drive is subverted by the backwards order of things behind the looking-glass, where the White Queen screams before she pricks herself, and to stay in one place you must run fast.

Alice in Wonderland

Tate Liverpool, UK.
4 November 2011
until 29 January 2012.
www.tate.org.uk/liverpool

NATURE.COM

For more on Charles Dodgson the mathematician: go.nature.com/fq9eem

Carroll seems to have conceived *Through the Looking-Glass* on the plane of a chess board, later adding the optical reversing effects of the mirror. That pairing echoes the thought experiment described by his friend, J. J. Sylvester, in his presidential address to the 1866 British Association for the Advancement of Science mathematical and physical section. Using the analogy of a flat bookworm on a flat page entering three dimensions when the page is curved, Sylvester said that our three-dimensional existence can reach into the fourth dimension in the same way,

“The Hatter has quarrelled with Time and now they are stuck: ‘It’s always six o’clock ...’”

“analogous to the rumpling of the page”. Alice, like the bookworm, can both move across the two-dimensional chessboard and bulge into a

different dimension through the mirror.

This suggests a new equivocal understanding of how time and space may be rumpled. In *Through the Looking-Glass*, Alice becomes aware that our mode of living in time is not the only pattern available. The Red Queen, hearing that days come one at a time in Alice’s country, says, “That’s a poor thin way of doing things. Now here, we mostly have days and nights two or three at a time, and sometimes in winter we take as many as five nights together — for warmth, you know.”

This merging and crossing between different modalities of time explains the attraction of Alice for the surrealists: Salvador Dalí’s brilliant illustrations pithily express the dream-time of Alice. They show the child leaping and dancing, with her shifting shadow always at just the wrong angle to the Sun.

Some of the more remarkable effects in



Alice’s trip through the looking-glass could have been inspired by the theory of other dimensions.

Alice are customary to us now. The elision and flow from one scene into the next (queen into sheep, shop into river) correspond as much to the editing processes of cinema as to the motions of dreams. Our familiarity with slow-motion photography may also make Alice’s leisurely fall into the underworld less astonishing, even though the alternatives suggested by the narrator offer the deep absurdity of the choice that is not a choice: “Either the well was very deep, or she fell very slowly, for she had plenty of time as she went down to look about her.” A body falling down a well, however deep, is rarely leisurely.

Carroll allows objects to retain their lethal weight in motion even while Alice floats down: gravity here is erratic in action. Alice must be ten times as heavy as the empty marmalade jar she picks off a shelf, but in

this instance she dawdles through the air while worrying about the jar falling and killing someone below. This reverses the usual error, mocked by philosopher John Stuart Mill — ‘the belief that a body ten times as heavy as another falls ten times as fast’.

Laws of motion had, for the Victorians, become one of the most controversial aspects of time. The Mad Hatter’s tea party combines the two. The Hatter claims Time as an ally during the tea-table argument, when Alice becomes exasperated, saying that the Hatter might do “something better with the time ... than wasting it in asking riddles that have no answers”.

“If you knew Time as well as I do,” said the Hatter, “you wouldn’t talk about wasting it. It’s *him*.” But it turns out that the Hatter has quarrelled with Time and now they are stuck: “It’s always six o’clock now ... it’s always tea-time, and we’ve no time to wash the things between whiles.”

Instead of time moving, they must move round the table forever as if on a clock face — and Alice soon ends up with the March Hare’s dirty tea-things in front of her. Tea-time is, of course, not an instant but a period, so the participants can continue their own lives and conversations within the arrested time. That the Hatter’s watch “tells the day of the month, and doesn’t tell what o’clock it is” is in keeping with this temporal stasis.

Only a few years after the Alice books, in 1874, German mathematician Georg Cantor argued in his theory of sets that there are degrees of infinity as well as an infinite infinity that, for mathematicians, eased the paradox of the continuity or discontinuity of motion. The teasing question of infinities had already been recognized by Alice who, talking to the Hatter about the process round the tea table, asks, “But what happens when you come to the beginning again?”

Luckily, Alice can walk away. She is not imprisoned in their eternal loop (and, it turns out, neither are they: the Hatter becomes a witness in the trial scene, and reappears in *Through the Looking-Glass* as the Anglo-Saxon messenger Hatta).

This is not a systematic fiction. It is a field of play. Time here, as in a mathematical manifold, makes Euclidean sense only locally; the whole resists resolution. The various forms of time will not lie still together; they are rumpled and energetic, endlessly alluring Alice. ■

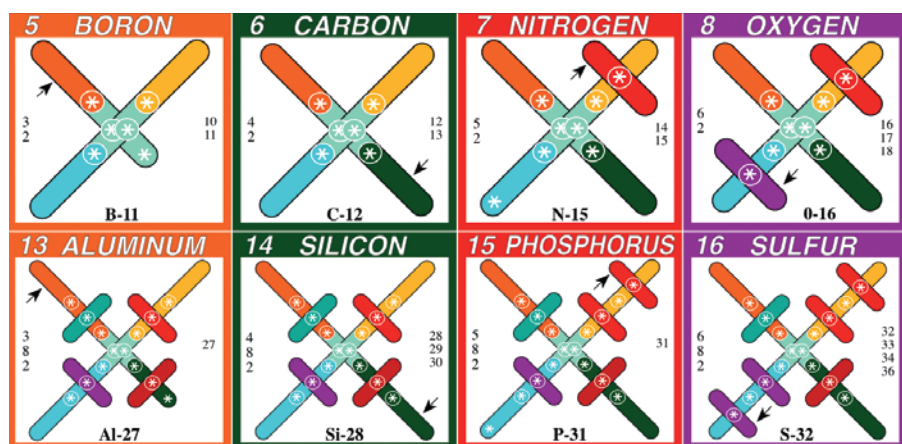
Gillian Beer is professor emeritus of English literature at the University of Cambridge, UK, a fellow of the British Academy and author of *Darwin’s Plots* (2009, third edition).

This is an edited and abridged version of *Time’s Manifolds* by Gillian Beer, taken from the catalogue accompanying the exhibition *Alice in Wonderland* at Tate Liverpool (Tate Publishing, 2011). Reproduced by permission of Tate Trustees, © Tate 2011.



The Mad Hatter’s non-stop tea party draws on ideas about infinity and the laws of motion.

J. CARTER



An interlocking 'circlon' model of elements is one idea developed by fringe theorist Jim Carter.

Q&A Margaret Wertheim

The outsider insider

Science writer Margaret Wertheim's latest book focuses on 'outsider physicists' — fringe theorists who probe the cosmos in their own way. On its publication, to be accompanied by a December exhibition at the newly opened Institute For Figuring gallery in Los Angeles, Wertheim explains her fascination with those who explore beyond the textbooks.

C. ALLAN



How did you get interested in fringe theories?

When I was a physics student, a professor told me that he had been getting strange scientific manifestos in the mail. I was a bit affronted that someone with little or no training in science would have the nerve to come up with a theory of everything. Years later, I wrote a column about a fringe physicist, and since then outsiders have sought me out. Almost all of them are interested in the fundamental nature of reality, and they are often outraged that physics is couched in abstract mathematical terms. They feel that physics has been hijacked, and that nature must speak a language that ordinary people can grasp. I've come to see them as the scientific equivalent of outsider artists. I want to understand what this phenomenon means in the context of our society.

What do your book and exhibition cover?

Physics on the Fringe looks at outsider physics by exploring the story of Jim Carter, the Leonardo da Vinci of fringe theorists. It then asks wider questions about the role of theoretical physics in the imaginative landscape of our culture. The first show at the new Los Angeles gallery of the Institute For Figuring — which I founded with my twin sister Christine to showcase the aesthetic and poetic dimensions of science and mathematics — will present the

Physics on the Fringe: Smoke Rings, Circlons, and Alternative Theories of Everything

MARGARET WERTHEIM

Walker: 2011. 336 pp. \$27

Physics on the Fringe Exhibition

Institute for Figuring, Los Angeles, California.

17 December 2011 until 30 March 2012.

work of Carter and other outsiders through diagrams, models and animations.

Who are these physics outsiders?

Quite a few are engineers, but they come from all walks of life. One is a retired California supreme court judge, another is a backyard car salesman. They have their own association, the Natural Philosophy Alliance, with a database listing annual meetings, published proceedings and articles by more than 2,000 theorists. One outsider I've met is a famous Hollywood film editor. He is obsessed with an eighteenth-century equation called Bode's law, which supposedly describes planetary orbits. There's a well-known Russian engineer who claimed to have invented a gravity-reducing device. And one Dutch theorist has proposed that the Universe is "a bouncing machine" shaped like a giant "twelve-lobed raspberry" that spews particles from a "Giant Virgin Black Hole".

What drew you to Jim Carter, the star of *Physics on the Fringe*?

One of his books landed in my hands in 1993. He had a sense of humour and I was

captivated by his intricate diagrams showing the ring-shaped particles, or 'circlons', that he believes compose all matter. His DIY approach to particle physics is part of a wider philosophy. He fixes his own cars and has built his own house, complete with a secret cave. He owns a company that makes devices for lifting sunken boats, and has gone hunting for a giant meteorite. It is not surprising that he has his own theories of the Universe too. His total physical and intellectual vision of science is antiquated but beautiful. It makes me wistful for the gentleman scientists of the nineteenth century.

Why listen to outsider physicists?

They may not be Albert Einstein or Paul Dirac. Their ideas aren't going to be taught at universities such as Princeton and Harvard. But their theories are a store of imaginative thinking about how our Universe might be constructed. This is an important cultural phenomenon — like studying the diaries of foot soldiers alongside those of generals. Some of the works, especially Carter's, are also aesthetic achievements. Most of all, they give us a window on to the role of science in our lives. These people want to be at home in the Universe. They believe that science can provide us with an understanding of the cosmos, but feel alienated by mainstream theories.

Why do they feel alienated?

The forefront of physics is inaccessible to most people. We have done the obvious stuff, and now we are in strange territory. What is physics for? In one sense, it is for making stuff work, like lasers and microchips. But this is not the whole answer. Johannes Kepler, Isaac Newton and Einstein wanted to comprehend the Universe and our place in it. Today's cosmological explanations have become incomprehensible to many people. This is one reason why religious fundamentalism has become reactive to science. If mainstream science ceases to provide us with an accessible picture of our world, it is not surprising that some folks begin to look elsewhere.

Can outsiders help to advance science?

There isn't a bright burning mark in the sand between truth and fiction. Things that seem fantastical in one era can become mundane in another. Concepts that were assumed to be true in the past, such as the 'element' phlogiston, now seem ridiculous. The idea that science proceeds purely by empirical scrutiny is a myth. Quantum mechanics makes predictions about matter and light that have been verified to many decimal places. Special relativity is used to correct for infinitesimal deviations in the position of Global Positioning System satellites. But how will our concepts fare in 500 years? We don't know. ■

INTERVIEW BY JASCHA HOFFMAN

Correspondence

European stem-cell ruling is misleading

Last month the European Court of Justice ruled that inventions derived from human embryonic stem cells are largely unpatentable. This ruling will shape the development of stem-cell technology. So, to prevent confusion, we wish to point out that the ruling contains crucial errors with respect to the underlying science.

At issue is the dividing line between what does and does not constitute an individual human. Under European law, individuals are not patentable. The new ruling misleadingly classifies pseudo-fertilized eggs, or parthenotes (lumping together those made with and without nuclear transfer), as requiring the protections of personhood. This classification was made on the grounds that these eggs are “capable of commencing the process of development of a human being just as an embryo created by fertilization of an ovum can”.

However, mammalian parthenogenetic embryos do not develop in the same way as normal embryos; nor are they developmentally viable if made without nuclear transplant. They are far down the sliding scale of developmental potential shared by all cells (including cells in the adult human body). It is therefore misguided for the ruling to put all parthenote-derived cells, which have technological potential (see A. A. Kiessling *Nature* **434**, 145; 2005), at the same end of the scale as cells that can fully differentiate (totipotent cells).

The ruling sensibly prevents a second confusion: inventions derived from human embryonic stem cells that require prior destruction of embryos are, it says, unpatentable. Because the technology already exists to make human embryonic stem-cell lines that preserve the viability of the donor embryo (Y. Chung *et al. Cell Stem Cell* **2**, 113–117; 2008),

embryo destruction is unnecessary. The scope of the ruling may therefore be narrower than some might conclude.

Whether one is for or against biological patenting (and we have no settled view or financial interest at stake here), it will be unfortunate if the wording of the European Court’s ruling should inadvertently inhibit a potentially useful, ethical technology using parthenotes just because of a lapse in scientific understanding.

Jeremy B. A. Green on behalf of 6 co-authors*, King’s College London, UK.
jeremy.green@kcl.ac.uk
Competing interests declared. See <http://dx.doi.org/10.1038/479041a> for declaration and *full author list.

Call to support Greek research reforms

As members of the Greek National Council for Research and Technology, we appeal to the global scientific community to lend us its support in Greece’s present critical economic situation. We must convince governments and relevant international bodies that we need a more realistic time frame to enable Greece to achieve the objectives that we all hold in common.

The council’s role is to advise on the reorganization of Greek research as an engine for future growth and as a way out of the current crisis. There is a grave danger, however, that the research infrastructure, and many institutions essential to its upgrading, could disintegrate because of the pressures being placed on the country.

Greece’s international creditors are demanding rapid restructuring of the entire research establishment, forcing financial cuts to be made indiscriminately, with serious implications for recovery prospects.

In working to reform the

research infrastructure, the council has seen that key stakeholders are determined to put Greece on a more effective track and are enthusiastic about radical change. The Greek scientific community contains pockets of excellence and much talent and dedication. It is imperative that Greece is allowed time to implement the much-needed reforms in a careful and orderly way.

Stamatios M. Krimigis Academy of Athens, Greece.

Aristides A. N. Patrinos Synthetic Genomics, La Jolla, California, USA.
apatrin@syntheticgenomics.com

Easter Island’s complex history

The history of Easter Island is important because it offers a lesson in long-term survival in an isolated and resource-poor environment. In his review of our book, *The Statues That Walked*, Paul Bahn makes some potentially misleading assertions (*Nature* **476**, 150–151; 2011).

He implies that we have overlooked other people’s work, but those studies are unpublished. We instead acquired our own data, controlling for content and quality.

Our excavations and radiocarbon dating indicate that Easter Island was colonized several centuries later than Bahn contends (T. L. Hunt and C. P. Lipo *Science* **311**, 1603–1606; 2006). Settlement of this and other islands in eastern Polynesia occurred over the past 800–1,000 years. On the basis of new evidence, most archaeologists working on Easter Island now reject the notion that its population collapsed before the arrival of Europeans.

It is generally agreed that the island was almost completely deforested by the time Europeans arrived in 1722. We never argued that rats were the only cause of deforestation, which happened over centuries and resulted

from people burning vegetation for agriculture, and from rat predation of seeds.

Although there is skeletal evidence for some violence on the island, only a few examples indicate mortal wounding. As we explain in our book, the statues were a focus of competitive signalling that staved off lethal violence.

Terry L. Hunt University of Hawaii-Manoa, Honolulu, Hawaii, USA.
thunt@hawaii.edu

Carl P. Lipo California State University, Long Beach, California, USA.

Control of dengue fever in Pakistan

Five years ago the Indian subcontinent experienced its first epidemic of dengue fever, with more than 5,000 people in India and more than 2,000 in Pakistan hospitalized (M. A. Rai and H. Khan *J. Clin. Virol.* **38**, 269–270; 2007). Pakistan is now in the middle of a dengue resurgence: more than 15,000 cases have been recorded in Lahore alone. This is potentially disastrous for the country’s health-care system, which is already on its knees.

Mortality is much higher this time. There have also been reports in Pakistan of resistance to agrochemicals evolving in dengue mosquitoes (H. A. Khan *et al. Parasite Vectors* **4**, 146; 2011), calling into question the current massive fumigation drive. Last year’s record floods have aggravated the situation.

Dengue experts have been flown in from Sri Lanka, Indonesia and the World Health Organization, but they may be too late. Prior planning and policy formulation by Pakistan’s health-care authorities are key to the prevention of future dengue outbreaks.

Mohammad A. Rai Nuffield Department of Medicine, University of Oxford, UK.
mohammad.raai@ndm.ox.ac.uk

Steve Jobs

(1955–2011)

The exacting visionary who put the personal into computing.

Apple co-founder Steve Jobs branded the Macintosh as “the computer for the rest of us”. Today, that includes toddlers too young to speak, who use Apple’s iPad without adult help. His uncompromising vision survived rejection, ejection from the company he founded, and a fairy-tale return that transformed Apple into the most valuable company in the world. His passion to make computers beautiful and accessible to everyone reshaped the personal computer industry — and also music, mobile phones, books, magazines and films.

Jobs, who died on 5 October at the age of 56, after a long battle with pancreatic cancer, framed the arc of the personal computer industry. Apple Computer, which he co-founded in 1976 with Steve Wozniak, was one of the first companies to build and sell computers for ordinary people; Apple’s iPhone and iPad arguably mark the transition to the ‘post-PC era’ of portable devices accessing cloud-computing applications.

Jobs was born in San Francisco, California. Adopted by Paul and Clara Jobs, whom he thereafter considered his true parents, he was raised in Cupertino, the heart of what was to become Silicon Valley. His father, a machinist, taught him electronics, and he attended after-school lectures at nearby Hewlett-Packard. Jobs later worked there as a summer employee, and from there lured Steve Wozniak to launch Apple Computer in 1976.

Like Bill Gates, Jobs was a college dropout. He attended Reed College in Portland, Oregon, for a single semester. His formative educational experiences were self-directed: his learning about phone hacking, his participation in the hobbyist Homebrew Computer Club, his travel to India and experiments with LSD. In the early days of a new field, it is often the autodidact and enthusiast who sees with fresh eyes and grasps opportunity.

Jobs had big dreams. In 1983, he invited John Sculley, the president of Pepsi-Cola, to serve as Apple’s chief executive with the challenge, “Do you want to sell sugar water for the rest of your life, or do you want to come with me and change the world?”

After he was ousted from Apple in 1985 by Sculley, Jobs founded NeXT (later NeXT Computer), a high-end workstation company, featuring an easy-to-use graphical

programming environment called NeXTSTEP. NeXT might have been a footnote in computer history but for two things: it was on a NeXT Computer that British computer scientist Tim Berners-Lee created the World Wide Web; and it was Apple’s purchase of NeXT in 1996 that brought Jobs back to the company. He served as interim chief executive of Apple from 1997, taking the title in full in 2000, and holding it until he resigned six weeks before his death.

In 1986, Jobs had acquired a computer



graphics company from filmmaker George Lucas, which he renamed Pixar. After struggling for years, it pivoted, and moved into computer-animated films. Pixar became famous in 1995 with *Toy Story*, the first full-length feature film made entirely with computer-generated imagery. Pixar went on to win 26 Academy Awards, 7 Golden Globes and 3 Grammy Awards, while turning animated films into a multibillion-dollar film category. In addition to being the majority owner and chief executive of Pixar, Jobs was credited as a producer on *Toy Story*.

Jobs’s career was marked by a series of milestone products. The Apple II, one of the first mass-produced computers, was released in 1977, four years before the IBM PC. The Macintosh, released in 1984, popularized the mouse-and-window paradigm for computing that had been developed at Xerox Parc. Mac OS X, the revamped Macintosh operating system based on NeXT, was introduced

in 2001, and led the resurgence of Apple. The same year welcomed the iPod — the personal music player that created a legitimate market for digital music and, through iTunes, made Apple the largest music retailer in the world.

The iPhone transformed the mobile-phone category on its release in 2007, and also demonstrated the first practical touch-screen computer interface; the iPad, released just last year, was the first successful mass-market tablet computer, so striking in its impact that it is almost impossible to remember a time before it existed. Each product was revolutionary in its design and had an effect both on the computer industry and on the way media are created and consumed. Jobs’s leadership saw Apple demonstrate an ability to create and reshape industries that brought him comparisons to Thomas Edison.

As a business leader, Jobs was famously demanding. He could be abrasive. At an executive retreat I attended, which was run by a large media company, he complained about the performance of the company’s iPad application, then remarked: “Of course, if your engineers were any good, they’d be working for us.” But his abrasiveness was also his genius. Like Michelangelo, he cut away what was unnecessary in pursuit of an uncompromising vision.

Beauty is not often considered in industry — or in science — and yet it was aesthetics that drove Jobs to demand excellence from his engineers. He knew what he wanted, and he wouldn’t stop until he achieved his vision. He was passionate about creating beautiful products that would delight customers.

Aesthetics was also the key to Jobs’s marketing genius. US art critic Dave Hickey once said that when products become commodities, as personal computers had become, they become “art markets” in which products are sold not on the basis of what they do but what they mean.

It is perhaps for this reason that Jobs was one of the rarest of breeds: a business leader who was loved. After his death, fans left flowers and notes at Apple stores around the world in an orgy of public grief reminiscent of that following the death of John Lennon or Princess Diana. In the words of W. H. Auden, on that day, “he became his admirers”. ■

Tim O’Reilly is chief executive of O’Reilly Media, Sebastopol, California 95472, USA.
e-mail: timoreilly@gmail.com

I. WALDIE/GETTY IMAGES

FORUM Societal collapse

Drought and the Maya

The collapse of the Maya civilization is often attributed to drought, but is the explanation really as simple as that? On the basis of evidence from their respective fields, an archaeologist and a palaeoclimatologist call for a more nuanced assessment.

THE TOPIC IN BRIEF

- The Mesoamerican Maya civilization was one of the most successful on the planet, reaching the peak of its development from about AD 250 to 750.
- In the period known as the Terminal Classic (roughly AD 750 to 1050), the Maya civilization underwent a profound crisis that

led to the abandonment of many sites.

- Palaeoclimate data have been used to argue that severe drought during the Terminal Classic may have caused this collapse.
- This argument, however, does not explain the complexity of the archaeological data, and is a matter of debate.

The story of the artefacts

JAMES AIMERS

For more than a century, the story of the Maya civilization has provided a narrative of precipitous rise followed by spectacular decline. Explanations offered for the demise of Maya civilization during the Terminal Classic period have varied in accordance with current events¹, ranging from war or political revolution to deforestation and climate change. Drought currently takes centre stage in many discussions. Some researchers espouse the extreme view that several crises throughout Maya history were caused by drought², whereas others offer more measured arguments. But the net result is that droughts are now routinely invoked^{3,4} to explain a supposedly pan-Maya collapse.

As David Hodell will argue (see later), palaeoclimate data in support of the drought hypothesis are not always unambiguous, and must be interpreted with care. Tidy tales of Maya collapse are also countered by a complex narrative that has emerged from archaeological research. Evidence for the collapse of Maya sites runs the gamut from precisely dated monuments of conquest to evidence from human bone, fauna and flora (which can also be directly dated). The combination of many archaeological data sets over decades of research has made it clear that sites in a range of physical environments underwent a variety of changes in the Terminal Classic — some much more dramatic than others — rather than marching synchronously towards oblivion.



Figure 1 | Later than expected. The decline of Maya civilization was once thought to have occurred between AD 830 and 900, but some sites, such as Lamanai, Belize (shown), were not abandoned until the seventeenth century.

Although the Terminal Classic period was first dated to AD 830 to 900, we now know that the changes associated with it occurred at different times and rates across the Maya lowlands (Fig. 1). Sites in the Petexbatún region of Guatemala were abandoned as early as the mid-eighteenth century; others, such as Chichén Itzá in Mexico, were not deserted until the mid-eleventh century (Fig. 2). In the Mopan

Valley of Guatemala, the Terminal Classic may have extended into the thirteenth century. Most surprising of all, sites such as Lamanai and Tipu in Belize, and sites in the Petén Lakes region of Guatemala, were not abandoned until well into the historic period — as late as 1697 for Tayasal in the Petén Lakes district.

Collapse may therefore not be the correct term to describe the variable, long-term process of Maya decline⁵, and some archaeologists suspect that climate scientists emphasize droughts that correlate with known abandonments, but ignore those that occurred in periods of growth. In fact, Mesoamerican civilizations in general experienced multiple periods of growth and decline — the Maya were far from singular in this respect.

No one doubts that devastating droughts occurred in the Maya lowlands in the past, as they do now. But so too did long and destructive wars, which show direct links to some site abandonments⁶. The escalating economic and political inequities of the Terminal Classic may also have had a role — perhaps after the 'Arab Spring' of 2011 we will see revolution revisited as an explanation for the Maya collapse. Drought has been convincingly shown to have been an important cause of abandonment at only a handful of sites; in other cases, archaeologists have looked carefully for evidence, to no avail.

Despite its popular appeal, drought as an explanation for the Maya collapse flattens the complexity of the archaeological record and ignores the ability of the Maya to react and adapt. Archaeologists have been too willing to overlook the problems inherent in the interpretation of palaeoclimate data, whereas climate scientists have too often relied on obsolete archaeological information. Part of the problem is that researchers from both fields write and present in different venues for different audiences. We must do a better job of collaborating if we are to be considered more than just good storytellers.

James Aimers is in the Department of Anthropology, State University of New York at Geneseo, Geneseo, New York 14454, USA. e-mail: aimers@geneseo.edu

B. HAAS/NATIONAL GEOGRAPHIC STOCK



Figure 2 | Dating decline and drought. Archaeological evidence (from sites marked in blue) suggests that the collapse of Maya civilization occurred at different rates at different places. Abandonments occurred at sites in the Petexbatún region in the eighth century, at Chichén Itzá in the eleventh century, and at sites in the Mopan Valley possibly as late as the thirteenth century. Lamanai, Tipu and some sites at the Petén Lakes were not abandoned until well into the historic period. Palaeoclimate evidence (from sites marked in red) tells the history of drought in Mesoamerica. Data from Lake Chichancanab show that

drought occurred between AD 800 and 1000, whereas evidence from Barranca de Amealco dates drought more precisely from AD 897 to 922. Other data suggest that a series of droughts occurred, from AD 760 to 910 (evidence from the Cariaco Basin) and from AD 800 to 950 (evidence from Tecoh). The disparity between the timing of site abandonments and periods of drought casts doubt on theories that drought caused a pan-Maya collapse. Palaeoclimate data must also be considered in context — changes in rainfall around the Cariaco Basin do not necessarily explain what happened 2,700 kilometres away at the Maya site of Tikal.

Maya megadrought?

DAVID HODELL

There are ample palaeoclimate data that support the drought hypothesis for Maya collapse, but its interpretation is not straightforward. Instrumental meteorological data for Mesoamerica are available for only about the past century. Documenting past droughts therefore relies on proxy indicators for rainfall, which are preserved in natural archives such as speleothems (stalagmites), tree rings and lake and marine sediment cores. Each archive and proxy has inherent strengths, but also weaknesses that contribute uncertainty to palaeoclimate inferences. For example, palaeoclimate records require a chronology, and issues of dating accuracy (the absolute age of a sample), temporal resolution (the finest interval of time that can be resolved) and stratigraphic correlation lie at the heart of reconstructing past climate from such data.

So what is the evidence for drought in the Terminal Classic? The first physical evidence⁷ came from measurements of oxygen isotopes in shells and gypsum found in sediment cores from Lake Chichancanab in the Yucatán Peninsula, Mexico (Fig. 2), which indicate that climate drying occurred between about AD 800 and 1000. Subsequent palaeoclimate studies have tried to define the nature and timing of this 'megadrought' more precisely. For example, variations in titanium concentrations in annual deposits contained in marine sediment cores from the Cariaco Basin, off

northern Venezuela, indicate that multi-year droughts occurred at about AD 760, 810, 860 and 910 (ref. 8). Oxygen-isotope data from an annually banded speleothem found in a cave at Tecoh (in Yucatán, Mexico) reveal eight severe droughts in the period from AD 800 to 950, each lasting from 3 to 18 years⁹. And most recently, a 1,238-year tree-ring record from Barranca de Amealco (Querétaro, Mexico) has provided evidence for a megadrought between AD 897 and 922, with shorter events centred at AD 810 and 860 (ref. 10).

But palaeoclimate records must be evaluated with respect to their location, chronology and the rainfall proxy used. For example, the farther an archive is from the Maya lowlands, the less confident one can be that a rainfall reconstruction applies to the Maya area. So, does the rainfall record from the Cariaco Basin⁸ really inform us about past precipitation at the Maya site at Tikal, Guatemala (abandoned about AD 900), some 2,700 kilometres away?

Another issue is the accuracy of dating droughts — annual resolution in a palaeoclimate record doesn't necessarily imply annual accuracy, because some chronologies are not anchored securely in time^{8,9} (although tree-ring chronologies provide both high resolution and accuracy¹⁰). Furthermore, the calibration between a proxy and rainfall is not always straightforward.

In summary, there is robust palaeoclimate evidence for drought during the Terminal Classic, especially in the ninth and early tenth centuries AD, but details remain vague. There was undoubtedly considerable regional and local variability in the timing and distribution of droughts in the Maya lowlands at this time. Difficulties arise when comparing

climate and archaeological records, because the data sets inadequately resolve spatial and temporal variability in climate and cultural systems, and both are dated with uncertainty¹¹. Our current understanding of the relationship between climate and Maya cultural change during the Terminal Classic therefore remains fuzzy. Sharpening the evidence will require new high-resolution, accurately dated records, preferably on a local scale, from both archaeologists and palaeoclimatologists, and mutual cooperation when interpreting the results. ■

David Hodell is in the Department of Earth Sciences, University of Cambridge, Cambridge CB1 1DS, UK.
e-mail: dah73@cam.ac.uk

1. Wilk, R. J. *Anthropol. Res.* **41**, 307–326 (1985).
2. Gill, R. B. *The Great Maya Droughts: Water, Life, and Death* (Univ. New Mexico Press, 2000).
3. Hunt, B. G. & Elliott, T. I. *Clim. Change* **69**, 393–407 (2005).
4. Cecil, J. *The Fall of the Mayan Civilisation* www.bbc.co.uk/history/ancient/cultures/maya_01.shtml (2011).
5. Aimers, J. J. *J. Archaeol. Res.* **15**, 329–377 (2007).
6. Inomata, T. in *The Archaeology of Settlement Abandonment in Middle America* (eds Inomata, T. & Webb, R. W.) 43–60 (Univ. Utah Press, 2003).
7. Hodell, D. A., Curtis, J. H. & Brenner, M. *Nature* **375**, 391–394 (1995).
8. Haug, G. H. *et al. Science* **299**, 1731–1735 (2003).
9. Medina-Elizalde, M. *et al. Earth Planet. Sci. Lett.* **298**, 255–262 (2010).
10. Stahle, D. W. *et al. Geophys. Res. Lett.* **38**, L05703 <http://dx.doi.org/10.1029/2010GL046472> (2011).
11. Yaeger, J. & Hodell, D. A. in *El Niño, Catastrophism, and Culture Change in Ancient America* (eds Sandweiss, D. H. & Quilter, J.) 187–242 (Harvard Univ. Press, 2008).

NEUROSCIENCE

Periodicity without rhythmicity

Grid cells confer a spatial impression of an animal's environment on the brain. Their firing patterns in a cave-dwelling bat reopen old questions about how they do this, and pose some compelling new ones. [SEE LETTER P.103](#)

LAURA LEE COLGIN

Grid cells are neurons that create a map of an animal's surroundings in a part of the brain known as the medial entorhinal cortex¹. They achieve this by firing when the animal passes through discrete, regularly spaced locations in the environment. These spatially periodic activity patterns are reminiscent of the temporally periodic theta-frequency waves (roughly 4–10 hertz) generated in the cortical electroencephalogram during spatial navigation — tempting an inference that theta rhythms underlie grid-cell responses. But on page 103 of this issue, Yartsev *et al.*² challenge this idea by showing that, in the absence of theta rhythms, grid cells still fire in bats.

Grid cells, which were first discovered in rats¹, give rise to distinctive geometric patterns of activity as an animal travels across locations that can be depicted as the vertices of a grid of equilateral triangles (Fig. 1). Two conflicting theories attempt to explain the neural mechanism that underlies these spatially periodic firing patterns of grid cells: attractor network models^{3–5} and oscillatory interference models^{6,7}.

Attractor network models propose that extensive recurrent connections among cells in the network drive the formation of grid-cell patterns in the absence of theta-frequency oscillations. Oscillatory interference models postulate that periodic grid-cell activity results from interference between different oscillators in the theta-frequency range in individual grid cells. Until now, it has not been possible to distinguish between these models experimentally, because rodent grid-cell activity coexists with continuous theta oscillations. By using a different mammalian species that lacks continuous theta oscillations — the Egyptian fruit bat — Yartsev *et al.*² avoid this problem. Their findings argue against the oscillatory interference models.

But how can grid-cell periodicity be maintained in the absence of theta oscillations? The answer may lie not in

the oscillatory activities of single cells but in interactions between multiple cells. However, the highest proportion of grid cells is found in layer II of the medial entorhinal cortex⁸, where excitatory recurrent connections are sparse or non-existent⁹. One solution to this seeming paradox is provided by multiple recordings from individual cells, which show that recurrent connectivity in layer II of the medial entorhinal cortex is primarily inhibitory¹⁰. In line with this, inhibitory intrinsic connectivity is sufficient to produce grid-cell activity patterns in at least one attractor network model⁵.

It was recently discovered that, in rats, grid-cell spatial periodicity is lost as a result of inactivation of a brain region called the medial septum^{11,12}. Because medial septal inactivation suppresses theta oscillations, these results seem to support the hypothesis that theta oscillatory mechanisms underlie periodic firing in grid cells. Yartsev and colleagues'

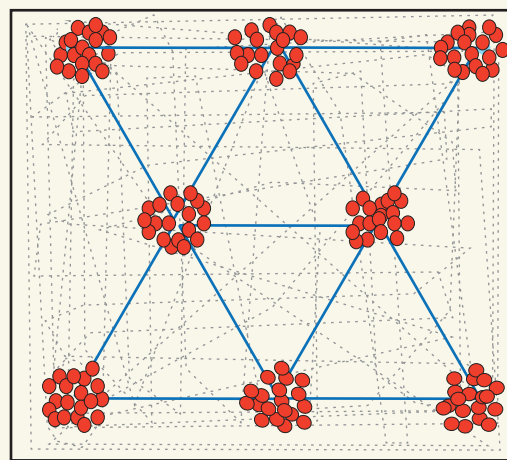


Figure 1 | Grid cells in rats and bats. Grid cells provide a neural representation of space, firing as an animal passes through regularly spaced locations (red circles) in its environment. In the mock patterns of grid-cell activation shown, grey lines track an animal's trajectory in a box enclosure. The activation patterns form a grid of equilateral triangles (blue lines), signifying that grid cells track precise distances and locations regardless of an animal's route. In rats, spatially periodic firing of grid cells is associated with characteristic temporal rhythms in the theta-frequency range. Yartsev *et al.*² show that in bats, grid-cell firing patterns are maintained in the absence of theta rhythms, a finding that undermines the idea that theta rhythms are relevant to grid-cell responses during navigation.

results, showing that bats do not need theta rhythms to maintain grid-cell firing patterns, challenge this interpretation and raise the question of why inactivation of the medial septum should abolish periodic activity patterns in grid cells in rats.

One answer could be that the loss of medial-septal input disturbs the balance between inhibition and excitation that is necessary to maintain grid-cell patterns in the neuronal networks of the medial entorhinal cortex. Another relates to the finding that the running speed of animals modulates the activity of medial-septal neurons¹³. Loss of velocity information from the medial septum might disrupt grid-cell firing, as both grid-cell models assume that velocity input is required to track an animal's spatial location and maintain grid-cell firing patterns.

Another explanation for why rats exhibit theta-rhythmic modulation of grid-cell firing¹⁴ but bats do not² may relate to the animals' different behaviours in the wild. Egyptian fruit bats travel back and forth from their home caves to locations where they found food on previous nights, suggesting that they encode spatial trajectories in memory¹⁵; however, this spatial memory encoding occurs during flying. Yartsev *et al.* recorded grid-cell activity in bats foraging under laboratory conditions — when they were crawling rather than flying. In the earlier experiments with rats^{11,12,14}, grid cells were recorded during running and walking, which are normal behaviours for these animals scavenging for food in their natural habitat.

Rats may automatically engage theta-associated memory-encoding machinery during running and walking because they need to remember where they could find food. By contrast, the brains of crawling bats may not need to be primed for spatial memory encoding, given that bats do not normally crawl while searching for food. It is possible that grid-cell activity in bats might be modulated by theta rhythms during more natural exploratory behaviour, or during tasks that require them to learn spatial trajectories.

Yartsev and colleagues' finding that bats maintain grid-cell firing patterns in the absence of theta rhythms also reopens an old question: what is the function of theta rhythms in the medial entorhinal cortex? In addition to suppressing theta rhythms in the medial entorhinal cortex¹⁶ and hippocampus¹⁷, damage to the medial septum also impairs spatial memory encoding¹⁶. Theta rhythms are therefore likely to be crucial in linking activity in the medial entorhinal cortex with activity in other brain areas (such as the hippocampus) during spatial-memory encoding.

And there are other questions. For instance, how do grid cells represent three-dimensional space during flying? Another puzzle is that in the event that certain behaviours cause bats' grid cells to become modulated by theta rhythms, do these cells fire across successive theta phases to encode sequences of spatial locations as rat grid cells do¹⁴? Regardless of the answers, bats have earned a place as essential participants in the formulation of future theories of grid-cell activity. ■

Laura Lee Colgin is in the Section of Neurobiology, Center for Learning and Memory, University of Texas at Austin,

Austin, Texas 78712, USA.
e-mail: colgin@mail.clm.utexas.edu

1. Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E. I. *Nature* **436**, 801–806 (2005).
2. Yartsev, M. M., Witter, M. P. & Ulanovsky, N. *Nature* **479**, 103–107 (2011).
3. Fuhs, M. C. & Touretzky, D. S. *J. Neurosci.* **26**, 4266–4276 (2006).
4. McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. & Moser, M.-B. *Nature Rev. Neurosci.* **7**, 663–678 (2006).
5. Burak, Y. & Fiete, I. R. *PLoS Comput. Biol.* **5**, e1000291 (2009).
6. Burgess, N., Barry, C. & O'Keefe, J. *Hippocampus* **17**, 801–812 (2007).
7. Hasselmo, M. E., Giocomo, L. M. & Zilli, E. A. *Hippocampus* **17**, 1252–1271 (2007).

8. Sargolini, F. *et al. Science* **312**, 758–762 (2006).
9. Dhillon, A. & Jones, R. S. G. *Neuroscience* **99**, 413–422 (2000).
10. Couey, J. J. & Witter, M. P. *Soc. Neurosci. Abstr.* no. 101.3 (2010).
11. Brandon, M. P. *et al. Science* **332**, 595–599 (2011).
12. Koenig, J., Linder, A. N., Leutgeb, J. K. & Leutgeb, S. *Science* **332**, 592–595 (2011).
13. King, C., Recce, M. & O'Keefe, J. *Eur. J. Neurosci.* **10**, 464–477 (1998).
14. Hafting, T., Fyhn, M., Bonnevie, T., Moser, M.-B. & Moser, E. I. *Nature* **453**, 1248–1252 (2008).
15. Tsoar, A. *et al. Proc. Natl Acad. Sci. USA* **108**, E718–E724 (2011).
16. Mitchell, S. J., Rawlins, J. N., Steward, O. & Olton, D. S. *J. Neurosci.* **2**, 292–302 (1982).
17. Rawlins, J. N., Feldon, J. & Gray, J. A. *Exp. Brain Res.* **37**, 49–63 (1979).

QUANTUM COMPUTING

Diamond and silicon converge

Diamond-based quantum computers could potentially operate at room temperature with optical interfacing, but their construction is challenging. Silicon carbide, used widely in electronics, may provide a solution. SEE LETTER P.84

ANDREW DZURAK

Quantum computers offer the prospect of dramatically outperforming conventional computers for difficult tasks such as code decryption — their processing speeds will often rise exponentially with the number of quantum bits (or qubits) of information involved^{1,2}. Diamond-based qubits³ have generated excitement because the time for which they can retain their logic state (the quantum coherence time) is much longer than the time it takes to perform qubit operations, even at room temperature. Furthermore, they can be read out by optical means, potentially allowing them to integrate with photonic quantum-communication systems. However, their manufacture is still embryonic and expensive, so the potential for expanding to a large-scale quantum processor using diamond remains uncertain. In this issue, Koehl *et al.*⁴ (page 84) describe qubits in silicon carbide that have remarkably similar properties to diamond-based qubits. Because silicon carbide is a semiconductor that is widely used in high-power electronics, this result improves the prospects for scaling up to large systems.

Silicon carbide is a compound of silicon and carbon that exists in many different crystal structures, or polytypes. It is often found as the hard, abrasive material carborundum. Its extensive use in high-power electronics (Fig. 1) is due to its high thermal conductivity, its ability to sustain high electric fields before breaking down, and its high maximum current

density. The latter two properties result from its large bandgap of about 3 electronvolts, compared with silicon's bandgap of around 1 electronvolt (the bandgap being the energy required to generate conduction electrons in the material).

Koehl and colleagues studied a polytype

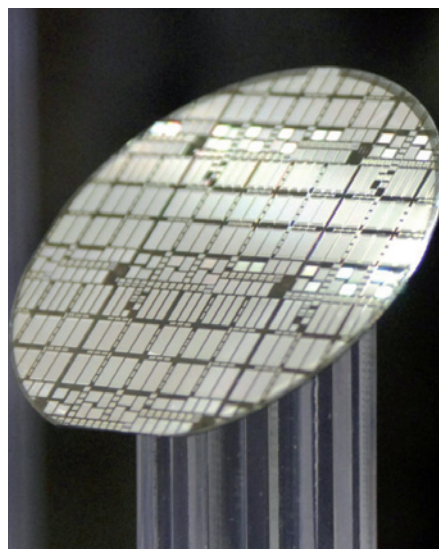


Figure 1 | Silicon carbide wafer. Koehl and colleagues' demonstration⁴ of spin qubits in silicon carbide — widely used for producing electronic devices such as those on this wafer — offers the tantalizing prospect of applying commercial manufacturing techniques to one day realize large-scale quantum computers using this technology.

of silicon carbide called 4H-SiC, which has a number of naturally occurring defects known as divacancies. These defects correspond to a missing silicon atom adjacent to a missing carbon atom in the crystal. They have much in common with a defect in diamond called the nitrogen-vacancy centre, which forms when a nitrogen impurity sits adjacent to a missing carbon atom in the diamond lattice. Both types of defect form a multi-electron system that has a net angular momentum (or spin) that can be used as a qubit. The authors⁴ show that several types of silicon carbide divacancy can act as optically accessible qubits, with some even having long coherence times at room temperature, just like those in diamond³.

A single electron in a static magnetic field is an ideal qubit because its spin orientation (and corresponding tiny magnetic moment) is aligned either parallel or antiparallel with the field, thereby encoding the two qubit logic states '0' or '1' used for computation. The diamond nitrogen-vacancy qubit is a little more complex, with several electrons associated with the crystal defect combining to create a total spin of 1 (in units of \hbar , the reduced Planck constant). This value leads to three possible spin states in a magnetic field (0 and ± 1), but two of these can be selected to form the qubit logic states. Divacancies in silicon carbide are also known^{5,6} to form multi-electron states with a spin of 1 that can be optically manipulated, and so demonstrating similar spin qubits in this material was a tantalizing prospect.

In their experiment, Koehl and colleagues⁴ measured the spin of the divacancies in 4H-SiC using an optical technique — photoluminescence — that is also used to measure diamond nitrogen-vacancy qubits. The technique involves directing a laser onto a sample and collecting the emitted fluorescent light. The fluorescence of the silicon carbide divacancies, like that of the diamond nitrogen vacancy, depends on their spin state, and so it is possible to read out the state of the qubits optically. Next, the authors borrowed another technique used to control diamond nitrogen-vacancy qubits³: by applying an oscillating

M. CARONNA/BLOOMBERG VIA GETTY

magnetic field at microwave frequencies, they performed electron spin resonance, in which the spin of a divacancy oscillates between its two qubit logic states. This type of control corresponds to performing a quantum 'write' operation.

Spin qubits are particularly attractive for solid-state devices, because they are only weakly disturbed by environmental electrical noise, leading to long coherence times. The situation is even better for qubits based on elements from group IV of the periodic table — such as carbon, silicon, germanium and their compounds, including silicon carbide — because they have a low background of nuclear spins that can destroy coherence, and in some cases they can be purified to remove the background spins almost completely. This is one reason why spin-based quantum-computing schemes using group IV materials^{3,7,8} have generated so much interest.

Furthermore, for diamond — and also now for silicon carbide, it seems — the reduced thermal vibrations that occur in these very hard, stiff materials means that the qubit states can have long coherence times (of around 100 microseconds or more), even at room temperature. Most other qubit technologies require very low temperatures to keep the qubit information intact, which is a drawback for a future involving portable quantum processors. However, it is most likely that early quantum computers will be fixed 'servers' that are accessed remotely, so their operating temperature will perhaps be less of an issue, at least for quite some time. Perhaps a more important feature of these silicon carbide qubits⁴ is their ability to emit light at wavelengths that can be transmitted using modern telecommunications infrastructure, offering the long-term prospect of connecting solid-state quantum computers within a secure quantum network.

Before such grand dreams can be realized, though, it will be necessary to construct a quantum processor that can outperform a conventional computer — and that will certainly take some effort. In modern-day microprocessor chips, a bit of information is stored and manipulated using silicon-based transistors, each less than 100 nanometres in size, with around one billion of them crammed onto a square centimetre of silicon. Although a quantum processor will not need as many qubits as a conventional processor to attack the same problem, it will still need a great many (probably many thousands) to run difficult algorithms such as code decryption, which involves factorizing very large numbers. Any serious qubit technology must therefore be manufacturable in a robust and precise manner. One advantage of silicon carbide over diamond is that integrated circuits that have many transistors are already routinely manufactured in this material⁹.

Although silicon carbide qubits offer

enticing prospects for quantum computing, a number of challenges for this new technology remain. First, the qubit operations reported by Koehl and colleagues⁴ were performed on a large ensemble of qubits, so the next step will be to demonstrate control and measurement of a single qubit. More significantly, technologies must be developed to 'engineer' thousands of individually addressable divacancy qubits, rather than merely identifying accidentally located defects. Engineering will also be needed to configure pairs of adjacent qubits reliably, to enable controlled two-qubit operations — another vital requirement for quantum computation. If these challenges can be met, silicon carbide could become a serious candidate for large-scale quantum computing. ■

IMMUNOLOGY

A heavyweight knocked out

Caspase-1 is one of the main culprits behind sepsis, a form of systemic inflammation. The related enzyme caspase-11 is also involved, but the relative roles of the two proteins have been confusing, until now. [SEE LETTER P.117](#)

DOUGLAS R. GREEN

Bacterial infections that cause sepsis are a problem of colossal proportions. In the United States alone, 18% of people admitted to hospital in 2000 as a result of this form of systemic inflammation died¹, and mortality associated with sepsis is rising. Studies of animal models have revealed that the molecular process underlying sepsis involves the proteolytic enzyme caspase-1. On page 117 of this issue, Kayagaki and colleagues² provide a crucial reappraisal of this conclusion. They show that mice lacking the gene that encodes caspase-1 also carry a mutation in a less-studied neighbouring caspase gene that is, in fact, responsible for some of the effects attributed to caspase-1 — including sensitivity to sepsis in animal models.

Caspase enzymes are perhaps best known for their function, in animals, in a type of programmed cell death called apoptosis³. Yet the first caspase to be discovered, caspase-1, was identified not in the context of apoptosis, but as an enzyme that processes the inflammatory mediator IL-1 β . There are 17 known caspases in mammals, and many others have been discovered in other classes and phyla. However, the functions of many of these enzymes are incompletely understood, and under-studied.

All caspases are expressed as inactive pro-enzymes. In mammals, three of these, the

Andrew Dzurak is at the Centre for Quantum Computation and Communication Technology, University of New South Wales, Sydney 2052, Australia.

e-mail: a.dzurak@unsw.edu.au

1. DiVincenzo, D. P. *Science* **270**, 255–261 (1995).
2. Bennett, C. H. & DiVincenzo, D. P. *Nature* **404**, 247–255 (2000).
3. Jelezko, F., Gaebel, T., Popa, I., Gruber, A. & Wrachtrup, J. *Phys. Rev. Lett.* **92**, 076401 (2004).
4. Koehl, W. F., Buckley, B. B., Heremans, F. J., Calusine, G. & Awschalom, D. D. *Nature* **479**, 84–87 (2011).
5. Baranov, P. G. *et al. JETP Lett.* **82**, 441–443 (2005).
6. Son, N. T. *et al. Phys. Rev. Lett.* **96**, 055501 (2006).
7. Kane, B. E. *Nature* **393**, 133–137 (1998).
8. Friesen, M. *et al. Phys. Rev. B* **67**, 121301(R) (2003).
9. Zetterling, C.-M. (ed.) *Process Technology for Silicon Carbide Devices* (Inst. Electr. Eng., 2002).

'executioner' caspases of apoptosis (caspase-3, -6 and -7), are present as inactive dimers that are activated by proteolytic cleavage^{3,4}. But cleavage does not activate all caspases. For most, including caspase-1, activation occurs when the proenzyme is recruited to a caspase-activation platform — a molecular complex specific to each caspase that brings caspase monomers close together to cause their assembly and activation. The significance of cleavage to caspase-1 activity remains uncertain, although it may stabilize the mature enzyme.

The activation platforms for caspase-1 are multiprotein complexes called inflammasomes⁵. These complexes often include the adapter molecule ASC, which is bound by several other proteins — such as NLR proteins — that sense different inflammatory triggers by assembling into inflammasomes (Fig. 1a). ASC also binds specifically to caspase-1, recruiting it to the inflammasome. On activation by an inflammasome, caspase-1 not only processes IL-1 β and its related pro-inflammatory protein IL-18, but also engages a secretory pathway for the release of these and other inflammatory mediators. Moreover, caspase-1 activation can cause cell death by the process of pyroptosis, which has only some features in common with apoptosis.

Undoubtedly, caspase-1 and its associated inflammasomes are important for the

response to a variety of pathogens and other inflammatory agents. For example, mice lacking caspase-1 are resistant to the blood toxicity (a model of sepsis) caused by injection of the bacterial surface molecule lipopolysaccharide⁶. But here things become a bit confusing. Ablation of another caspase, caspase-11, also protects mice from lipopolysaccharide toxicity⁷. Although in such mice caspase-1 activation is also compromised, it has been difficult to place caspase-11 into a scheme of caspase-1 activation. A study of human cells⁸ suggested that caspase-5 (the human version of the mouse caspase-11) assembles with NLRP1, ASC and caspase-1. However, subsequent studies of caspase-1 inflammasomes have not identified caspase-11 as part of the inflammasomes.

Kayagaki *et al.*² now provide some much-needed clarification. Strikingly, they show that caspase-11 is dysfunctional in strain 129 mice, which are widely used in studies involving genetic manipulation. Because the location of the caspase-1 gene in the genome is close to the gene encoding caspase-11, all caspase-1-deficient mice that have been generated in this strain are almost certainly also defective in caspase-11. Kayagaki and colleagues therefore generated new mice that are deficient in caspase-1, but not caspase-11, and show through elegant studies that, regardless of stimulus, caspase-1 is required for IL-1 β processing and secretion. Nonetheless, in some cases caspase-11 also seems to be required — for instance, for the response to cholera toxin B subunit, and to infections with the bacteria *Escherichia coli*, *Citrobacter rodentium* and *Vibrio cholerae*. These stimuli trigger cell death in a manner that depends on caspase-11, but not on caspase-1 or components of the caspase-1 inflammasome (Fig. 1b).

The activation platform for caspase-11 remains unknown, but does not seem to require ASC or other components of the caspase-1 inflammasomes². In addition, because cleavage per se does not activate caspase-1, a simple model in which caspase-11 cleaves caspase-1 is insufficient to explain its role in caspase-1 activation in response to these stimuli. It is possible that once the caspase-11 activation platform forms, caspase-11 recruits caspase-1 monomers to the complex, activating them through close proximity. Indeed, caspase-11 and caspase-1 seem to bind to each other, and it would be informative to determine whether the proteolytic activity of caspase-11 is required for caspase-1 activation after such interaction. Alternatively, it could be that caspase-1–caspase-11 dimers create an active enzyme, although currently there is no evidence for this.

Remarkably, Kayagaki and co-workers show that *in vivo* toxicity after lipopolysaccharide injection depends not on caspase-1 or components of the caspase-1 inflammasomes, but on caspase-11; caspase-1 seems to have

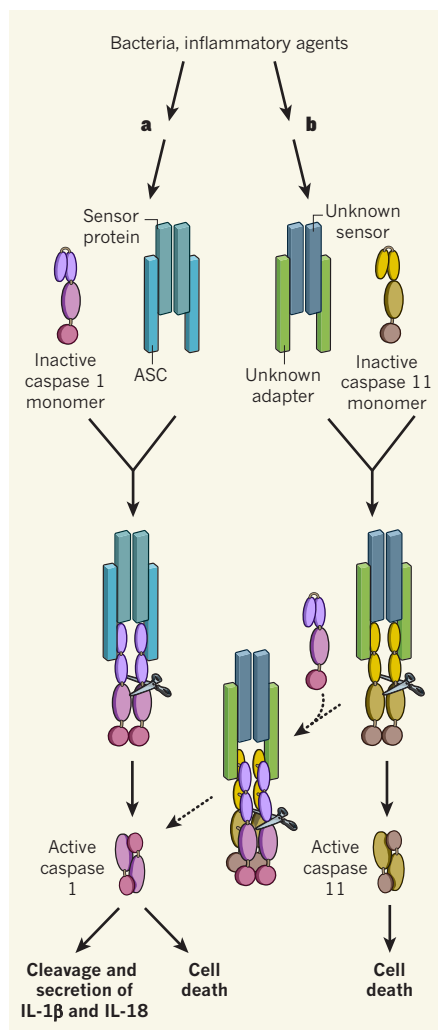


Figure 1 | Caspase-1 versus caspase-11.

a, Bacterial and inflammatory agents are known to activate different cellular sensor proteins. Through the adapter molecule ASC, these proteins recruit the inactive monomers of caspase-1, activating them by bringing them together; this promotes cleavage of caspase-1, which may stabilize the mature enzyme. Caspase-1 then processes IL-1 β and IL-18 and mediates their secretion and that of other inflammatory proteins; it can also trigger cell death, probably by cleaving and thereby activating executioner caspases. **b**, Kayagaki *et al.*² show that certain bacterial agents activate caspase-11 by an unknown mechanism. Caspase-11 can then promote cell death, but cannot process IL-1 β . Caspase-11 also promotes activation of caspase-1, most likely by recruiting caspase-1 to the caspase-11 activation platform (broken arrows).

an exacerbating function that is independent of roles for IL-1 β or IL-18. Although often used as a model of sepsis, lipopolysaccharide-induced toxicity does not show all the hallmarks of sepsis. So the functions of caspase-1 and caspase-11 should be explored in more relevant models.

Another ambiguity is how caspase-11 activation actually causes disease in response to lipopolysaccharides. IL-1 β and IL-18 are dispensable for the toxic effects, and whether

caspase-11 promotes non-canonical secretion is obscured by the cell death it causes — because cell death itself releases proteins non-specifically. Notably, activation of caspase-1 or caspase-11 led to cellular release of another inflammatory mediator, HMGB1, and neutralization of HMGB1 has protective effects against lipopolysaccharides *in vivo*⁹. In addition, another caspase, caspase-7, has a role in lipopolysaccharide-induced toxicity⁸. How caspase-11, but not caspase-1, fits into this scheme is not clear.

Finally, it seems remarkable that, despite having had access to the full genomic sequence of the 129 mouse strain for several years,

researchers are only now realizing that the gene encoding caspase-11 is dysfunctional in this strain. Even a recent analysis¹⁰ of the 129 strain — and of the C57BL/6 strain, which is also commonly used in laboratories — did not detect the caspase-11 mutation Kayagaki *et al.* describe (a five-nucleotide deletion), despite finding 24 deletions in genes that are expressed in the 129 strain, including several with immune functions. A general analysis that captures this particular mutation² may well reveal others. What else have we been missing? ■

Douglas R. Green is in the Department of Immunology, St Jude Children's Research

Hospital, Memphis, Tennessee 38105, USA.
e-mail: douglas.green@stjude.org

1. Martin, G. S. *et al.* *N. Engl. J. Med.* **348**, 1546–1554 (2003).
2. Kayagaki, N. *et al.* *Nature* **479**, 117–121 (2011).
3. Fuentes-Prior, P. & Salvesen, G. S. *Biochem. J.* **384**, 201–232 (2004).
4. Green, D. R. *Means to an End: Apoptosis and Other Cell Death Mechanisms* (Cold Spring Harbor Lab. Press, 2011).
5. Schroder, K. & Tschopp, J. *Cell* **140**, 821–832 (2010).
6. Li, P. *et al.* *Cell* **80**, 401–411 (1995).
7. Wang, S. *et al.* *Cell* **92**, 501–509 (1998).
8. Lamkanfi, M. *et al.* *Blood* **113**, 2742–2745 (2009).
9. Lamkanfi, M. *et al.* *J. Immunol.* **185**, 4385–4392 (2010).
10. Keane, T. M. *et al.* *Nature* **477**, 289–294 (2011).

system apart or prevent an already-formed cyclone from intensifying.

Typical shear conditions over the Arabian Sea are large enough that cyclones rarely develop into major events. This is especially true during boreal summer monsoon months, when warm surface waters extend into the North Indian Ocean. During these months, even with the warm surface temperatures, the monsoon's atmospheric circulation produces strong environmental shear that can suppress cyclone formation completely. However, an upswing in intense pre-monsoon tropical cyclones during the past 15 years is raising serious concerns that environmental change may be responsible for increasing cyclone activity, as opposed to these being a string of anomalous events. The observed increase may be part of a broad-scale shift in the regional climate, which could be directly attributable to anthropogenic emissions.

Over much of south Asia, anthropogenic air pollution has led to the formation of thick layers of haze known as atmospheric brown clouds⁵. The main sources of the pollution are fossil-fuel consumption and biomass burning, which deposit black carbon in the atmosphere, with serious negative consequences for human health⁶. This pollution produces brownish clouds of aerosol particles that can be several kilometres thick. The hazy conditions spread out over the Arabian Sea, blocking some of the Sun's energy and preventing it from reaching the sea surface, and thus causing cooling in the upper ocean.

Evan *et al.*³ propose that these brown clouds have an interesting effect on tropical cyclone environments. They argue that, by inhibiting the amount of incoming solar energy at the sea surface, brown clouds can effectively reduce the warmest Arabian Sea temperatures relative to the equatorial Indian Ocean. This smaller sea surface temperature difference between north and south could, in turn, reduce vertical wind shear during tropical cyclone seasons. Under this brown-cloud

CLIMATE CHANGE

Man-made cyclones

A recent surge in the intensity of tropical cyclones in the Arabian Sea has brought unprecedented damage and loss of life. Anthropogenic air pollution might be increasing the destructiveness of these storms. SEE LETTER P.94

RYAN L. SRIVER

A tropical cyclone is a rare occurrence. It requires a delicate combination of environmental conditions, including a particular amount of heat in the upper ocean, a certain distribution of moisture in the atmosphere, and a distinct structure of horizontal atmospheric winds. Although ocean temperatures in the Arabian Sea are warm enough for tropical cyclone development during much of the year, the region is relatively inactive, with only two or three events occurring annually. However, the occasional strong cyclone can wreak havoc in the region. In 1998, a major cyclone resulted in more than 1,100 deaths in western India, and Cyclone Gonu in 2007 caused more than US\$4 billion in collective damage to Oman, the United Arab Emirates and Iran¹ (Fig. 1). The frequency of intense cyclones such as these has risen in recent years, and there is evidence that the observed increase is the result of a general shift towards conditions more favourable for intense tropical cyclones². On page 94 of this issue, Evan and colleagues³ offer insight into this situation by examining the role of regional air pollution in creating these cyclone-friendly conditions.

It has long been known that the presence of strong vertical wind shear is a major factor in limiting tropical cyclone activity over the Arabian Sea⁴. Vertical wind shear is a measure of the difference in speed between the horizontal winds

measured near the sea surface and those high up in the atmosphere, in the outflow regions of a developing storm. Environments with strong shear will stretch a tropical cyclone by pushing the top of the storm away from the bottom. This shearing effect can tear a developing



Figure 1 | Aftermath of Cyclone Gonu. Evan and colleagues³ find that anthropogenic air pollution and global warming might be responsible for increasing the destructiveness of intense tropical cyclones such as Gonu, which occurred in the Arabian Sea in 2007 and devastated countries such as Iran (shown).

AFP/GETTY

scenario, tropical cyclones would intensify into much more powerful systems than would be possible in an environment without pollution (and with stronger shear).

Using observational evidence from the past 30 years, Evan and colleagues show that there is an observable increase in the average Arabian cyclone intensity between the periods 1979–96 and 1997–2010, which coincides with a reduction in vertical wind-shear conditions occurring within the storms. The authors attribute these changes to the presence of atmospheric brown clouds, supporting this hypothesis with accompanying climate-model simulations. Their findings point to a causal link between increasing air pollution over south Asia and more frequent intense tropical cyclone events, suggesting that human activity may play a direct part in modulating cyclone activity over the Arabian Sea.

The authors' results³ shed much-needed light on the relatively unexplored topic of Arabian Sea cyclones and climate. They imply that fascinating connections exist between tropical cyclone activity, air pollution in south Asia and global warming that have strong social and economic implications. Although intriguing, interpretation of these results must be treated with caution. For instance, the authors' analysis focuses on about 20 cyclones during the past 30 years, corresponding to events occurring before the annual onset of the monsoon. Out of these 20 events, the 5 most powerful storms occurred between 1998 and 2010, and it is this small fraction that determines the positive shift in cyclone intensity during the past 15 years. This is a very small number of events to use as a basis for estimating climate trends. Given the small sample size and handful of anomalous intense storms, it is difficult to disentangle meaningful climate signals from random variability of the sample.

Evan *et al.* hypothesize that the reduced shear observed in and around recent intense cyclones is indicative of a broader-scale negative trend. This proposal generally agrees with their model simulations and observational analysis. However, it is at odds with other recent work¹, which finds no discernible trend in Arabian Sea vertical shear within cyclone-development regions during the past 60 years. Perhaps surprisingly, most of the intense recent storms occurred during years exhibiting seasonal shear conditions that were either near or above the 60-year average.

Furthermore, if environmental shear over the Arabian Sea is decreasing, then should there not also be a change in the number of storms affecting this region? The potential for storm formation is quantifiable, because it depends primarily on the large-scale environmental factors previously discussed. Trends in this 'genesis potential'⁷ have been increasing in the Arabian Sea since 1980, but the average potential during the past 30 years

is substantially lower than that during the previous 30-year period¹. In other words, environmental conditions from 1950 to 1980, prior to the onset of major air pollution, were potentially more favourable for cyclone formation than the period from 1980 to 2010 (although the reliability of the early portion of the record may be limited by data quality). This suggests that changes in natural environmental factors could influence the variability of tropical cyclone activity in the Arabian Sea beyond what the authors attribute to anthropogenic brown clouds. Further work is clearly needed to understand the roles of natural climate variability and human-induced effects in the context of evolving data quality.

Limitations aside, Evan and colleagues' provocative findings³ raise fresh concerns about the potential hazards of anthropogenic aerosol pollution. If we continue on the path of unmitigated carbon emissions and brown clouds, then time will ultimately reveal whether the authors' results are robust. Nevertheless, the fact that these findings suggest the damaging

effects of brown clouds extend beyond known health impacts⁶ perhaps provides more urgency for instituting a sound strategy to minimize potential negative consequences. ■

Ryan L. Srivier is in the Department of Geosciences, Pennsylvania State University, University Park, Pennsylvania 16802, USA. e-mail: rsrivier@psu.edu

1. Evan, A. T. & Camargo, S. J. *J. Clim.* **24**, 140–158 (2011).
2. Rao, V. B., Ferreira, C. C., Franchito, S. H. & Ramakrishna, S. S. *V. S. Geophys. Res. Lett.* <http://dx.doi.org/10.1029/2008GL034729> (2008).
3. Evan, A. T., Kossin, J. P., Chung, C. E. & Ramanathan, V. *Nature* **479**, 94–97 (2011).
4. Gray, W. M. *Mon. Weath. Rev.* **96**, 669–700 (1968).
5. Ramanathan, V. *et al. Proc. Natl Acad. Sci. USA* **102**, 5326–5333 (2005).
6. United Nations Environment Programme *Integrated Assessment of Black Carbon and Tropospheric Ozone — Summary for Decision Makers* UNEP/GC.26/INF/20 (UNEP & World Meteorological Organization, 2011).
7. Camargo, S. J., Emanuel, K. A. & Sobel, A. H. *J. Clim.* **20**, 4819–4834 (2007).

PALAEONTOLOGY

Fresh light on southern early mammals

Little is known about mammalian evolution in South America during the age of the dinosaurs. The discovery of 100-million-year-old skulls confirms that mammalian faunas were endemic in southern continents at this time. [SEE LETTER P.98](#)

CHRISTIAN DE MUIZON

Mammalian remains from the age of the dinosaurs — the Mesozoic era, from 250 million years (Myr) to 65 Myr ago — are rare. Ten times fewer mammalian genera have been identified from the Mesozoic than from the age of mammals that followed it (the Cenozoic era, from roughly 65.5 Myr ago to the present), even though the Cenozoic has lasted less than half the time. Furthermore, Mesozoic mammals are most often known from isolated teeth or partial jaws; complete skulls and/or skeletons are exceptional. Our knowledge of the first two-thirds of mammalian evolution, which extends from the first record of a mammal about 220 Myr ago to the end of the Cretaceous period 65.5 Myr ago, is therefore terribly incomplete.

Considering the poverty of the fossil record, any discovery of a reasonably well-preserved skull of a Mesozoic mammal is a major palaeontological event. This is especially true in South America, where just one Mesozoic mammal is known from well-preserved skulls and skeletons, and the few other such taxa are

represented only by isolated teeth and jaws. On page 98 of this issue, Rougier *et al.*¹ describe the second Mesozoic mammal of South America to be represented by well-preserved skulls and jaws: *Cronopio dentiactutus*, found in Argentinian sedimentary rocks dating from the early Late Cretaceous (about 100 Myr ago), an epoch for which no mammals were previously known on the subcontinent.

Cronopio belongs to the superorder Dryolestoidae, whose members are regarded as close relatives of the modern therians (mammals that include the marsupials and placentals). Dryolestoid remains have been found mainly in the northern continents, but Rougier and colleagues' discovery, taken together with findings of teeth and jaws from 70-Myr-old deposits in Patagonia², reveals that an important evolutionary radiation of these mammals (an increase in taxonomic diversity) also occurred in the south during the Cretaceous.

The discovery of *Cronopio* is especially notable because it provides for the first time the whole cranial morphology of a dryolestoid. It also reveals some of the steps in the evolution



Figure 1 | South American mammals from the early Late Cretaceous. Rougier *et al.*¹ report the discovery in Argentina of two well-preserved, 100-million-year-old skulls from a previously unknown mammal, *Cronopio dentiactutus*. This artist's reconstruction reveals how the shrew-sized *Cronopio* might have looked. (Illustration by Jorge Gonzalez.)

of the skull that occurred during the transition of dryolestoids to therians. On the basis of tooth morphology, it was known that dryolestoids were taxonomically abundant and ecologically diversified in Patagonia during the Late Cretaceous^{2–4}, and that some genera extend into the early Tertiary⁵ (65.5 to 2.6 Myr ago). The new discovery¹ therefore extends the South American stratigraphic record of the group about 30 Myr further back.

So far, no marsupials or placentals have been securely identified in South America from the Mesozoic, even though they are abundant and well diversified at the very start of the Cenozoic^{6–8}. This pattern of mammalian evolution in South America apparently differs from that of the northern continents, where marsupials and placentals represented the major component of mammalian faunas during the Cretaceous. This in turn suggests that a major expansion and diversification of dryolestoids occurred in South America during the Cretaceous⁴. The fact that *Cronopio* and all the other known South American dryolestoids (except for one genus) are members of the same, apparently endemic taxonomic cluster, as shown by Rougier *et al.*¹, clearly corroborates this hypothesis. Moreover, the authors point out that even though *Cronopio* is much older than the other South American dryolestoids from the Late Cretaceous, it is already highly specialized, with an unusually long snout, dagger-like canines and a sophisticated masticatory musculature (Fig. 1). This further strengthens the idea of a major, endemic

evolutionary radiation of dryolestoids in South America during the Cretaceous, probably starting in the Early Cretaceous.

So how does *Cronopio* fit into the bigger picture of mammalian evolution in South America? Therians were probably absent from South America for most of the Cretaceous. North and South America were separated during the Early and most of the Late Cretaceous⁹, but a temporary continental bridge existed between them at the end of this period (between 75 and 65 Myr ago), which would have allowed therians to enter South America from the north. Indeed, the earliest therians of South America (the 64-Myr-old fauna from Tiupampa, Bolivia) have obvious close affinities with the North American faunas from the Late Cretaceous and the start of the Tertiary, as would be expected if the South American therians' ancestors had recently immigrated from the north^{7,8}.

Nevertheless, one must be careful when speculating about mammalian radiation in South America during the Mesozoic, no matter how tantalizing the evidence, as the fossil record is still desperately poor. Prior to the appearance of mammals in Patagonia in the Late Cretaceous², only two South American mammals from the Cretaceous have been identified: *Vincelestes neuquenianus*, from the Early Cretaceous (about 130 Myr ago), which is known from exquisitely preserved skulls and skeletons; and *Cronopio*, as now reported by Rougier and colleagues¹. But even these remarkably preserved fossils tell us almost

nothing about mammalian radiations in South America during most of the Cretaceous.

It is worth noting that therians were present during the Early Cretaceous in Africa¹⁰, which was still connected to South America at that time. It would therefore not be surprising if some early therians, or indeed any Early Cretaceous African mammal, appeared in the South American fossil record well before the Late Cretaceous. South America was part of Gondwana (an ancient 'supercontinent' that pre-dated the modern Southern Hemisphere landmasses) until about 100 Myr ago, which means that Cretaceous mammals could also have entered from there. Furthermore, monotremes — the group of egg-laying mammals that includes the modern platypus — have been recovered¹¹ in Patagonia from the early Tertiary, and monotreme-related mammals appear in the same region as early as about 160 Myr ago, during the Jurassic period¹². Some similar mammals could therefore very well have encompassed a major radiation during the Cretaceous in South America.

Rougier and colleagues' findings¹ clearly cannot be expected to solve the mystery of South American mammalian radiation on their own, but such discoveries of remarkably complete Mesozoic fossils always represent giant steps in the progress of mammalian palaeontology. In fact, one reasonably preserved Mesozoic mammalian skull in a critical stratigraphic and geographic position can be more relevant to our understanding of mammalian evolution and biogeography than hundreds of isolated teeth — even if teeth are the most common (and sometimes the only) remains that palaeontologists have to work with. ■

Christian de Muizon is in UMR 7207 CR2P (CNRS-MNHN-UPMC), Département Histoire de la Terre, Muséum National d'Histoire Naturelle, Paris 75005, France. e-mail: muizon@mnhn.fr

1. Rougier, G. W., Apesteguía, S. & Gaetano, L. C. *Nature* **479**, 98–102 (2011).
2. Bonaparte, J. F. & Migale, L. A. *Protomamíferos y mamíferos Mesozoicos de América del Sur* (Museo de Ciencias Naturales Carlos Ameghino, 2010).
3. Rougier, G. W., Forasiepi, A. M., Hill, R. V. & Novacek, M. *Acta Palaeontol. Pol.* **54**, 195–212 (2009).
4. Rougier, G. W., Chornogubsky, L., Casadio, S., Arango, N. P. & Giallombardo, A. *Cretac. Res.* **30**, 223–238 (2009).
5. Gelfo, J. N. & Pascual, R. *Geodiversitas* **23**, 369–379 (2001).
6. Gelfo, J. N., Goin, F. J., Woodburne, M. O. & de Muizon, C. *Palaeontology* **52**, 251–269 (2009).
7. de Muizon, C. & Céspedes-Paz, R. in *The Origins and Evolution of Cenozoic South American Mammals* (eds Rosenberger, A. L. & Tejedor, M. F.) (Springer, in the press).
8. Ladevèze, S., de Muizon, C., Beck, R. M. D., Germain, D. & Céspedes-Paz, R. *Nature* **474**, 83–86 (2011).
9. Iturralde-Vinent, M. A. *Int. Geol. Rev.* **48**, 791–827 (2006).
10. Sigogneau-Russell, D. *Bull. Mus. Natl. Hist. Nat. Paris* **16**, 291–312 (1995).
11. Rauhut, O. W. M., Martin, T., Ortiz-Jaureguizar, E. & Puerta, P. *Nature* **416**, 165–168 (2002).
12. Rougier, G. W., Martinelli, A. G., Forasiepi, A. M. & Novacek, M. J. *Am. Mus. Novit.* no. 3566 (2007).

Subsurface water and clay mineral formation during the early history of Mars

Bethany L. Ehlmann^{1†}, John F. Mustard², Scott L. Murchie³, Jean-Pierre Bibring¹, Alain Meunier⁴, Abigail A. Fraeman⁵ & Yves Langevin¹

Clay minerals, recently discovered to be widespread in Mars's Noachian terrains, indicate long-duration interaction between water and rock over 3.7 billion years ago. Analysis of how they formed should indicate what environmental conditions prevailed on early Mars. If clays formed near the surface by weathering, as is common on Earth, their presence would indicate past surface conditions warmer and wetter than at present. However, available data instead indicate substantial Martian clay formation by hydrothermal groundwater circulation and a Noachian rock record dominated by evidence of subsurface waters. Cold, arid conditions with only transient surface water may have characterized Mars's surface for over 4 billion years, since the early-Noachian period, and the longest-duration aqueous, potentially habitable environments may have been in the subsurface.

Chemical interactions between water and rock transform nominally anhydrous materials formed by volcanic and igneous processes into hydrous phases, including clay minerals¹ (phyllosilicates), which incorporate OH or H₂O in their structures (Table 1). On Earth, clay minerals form from near-surface weathering to produce soils, in hydrothermal systems located mostly at sea-floor spreading centres, or more rarely by direct precipitation in lake basins². Clays are then eroded, transported, buried and metamorphosed in processes driven by climate and tectonics³. The chemical changes accompanying clay formation influence global geochemical cycles⁴, and the types and distributions of clay minerals and their stratigraphic contexts record and permit reconstruction of environmental histories on global and regional scales^{3,4}.

The length of this terrestrial record is truncated, however. Plate tectonics has destroyed or obscured most of the record of Earth's first billion years, a critical period encompassing formation of the crust, heavy impact bombardment and the emergence of microbial life. In contrast, half of Mars's surface dates from its Noachian and Hesperian periods⁵ (4.1–3.7 billion years (Gyr) ago and 3.7–3.1 Gyr ago, respectively). The early Martian record is disrupted but preserved, and the composition and stratigraphy of units with alteration minerals record changes in water availability, climate and the nature of Mars's earliest aqueous, potentially habitable environments.

Globally widespread bedrock exposures of clay minerals (Fig. 1) indicate that long-term interaction with liquid water altered large portions of Mars's basaltic crust early in its history^{6–8}. Clay minerals are exposed in ancient, heavily cratered Noachian terrains and contrast markedly with younger terrains, which are dominated instead by crystalline igneous materials⁹ with weakly altered and possibly coated surfaces¹⁰ as well as by spatially restricted sulphate-bearing layered deposits^{7,11}. The apparent restriction of clays to Mars's oldest Noachian terrains and of sulphates largely to younger Hesperian terrains has been proposed to indicate a transition from an early, comparatively water-rich era with near-neutral-pH weathering to a more arid era with saline and acidic surface and near-surface waters^{7,12}. Data from *in situ* rover measurements, orbital hyperspectral imaging, geochemical models and hydrological models have converged to support formation of late-Noachian to Hesperian sulphate-bearing deposits by precipitation of

salts from evaporating, sometimes acidic, groundwater discharge¹³. However, earlier Noachian environmental conditions resulting in clay formation are less well understood. In particular, it is not known whether clay formation occurred mainly at the surface or in the subsurface⁷.

This issue has a substantial bearing on understanding Mars's early evolution. At present, the average atmospheric pressure on Mars is 7 mbar, the average temperature is 218 K, and ice typically sublimates to the water-undersaturated atmosphere without passing through the liquid phase¹⁴. However, valley networks, deltas and outflow channels provide geomorphic evidence of Noachian and Hesperian surface waters^{15,16}. These observations have led to the promulgation of atmospheric models involving carbon dioxide pressures 100–1,000 times higher than at present, or a distinctly different trace gas composition, to produce the higher pressures and temperatures needed to sustain surface liquid water and possibly precipitation^{17,18}. It has proven problematic for models to produce sufficient greenhouse warming and, moreover, to explain what happened to the proposed thick early atmosphere. Loss to space and trapping in carbonate rocks, ice or clathrates have been proposed, although current estimates are that these loss fluxes are insufficient and that trapping reservoirs hold only a few tens of millibars or less of carbon dioxide¹⁹. Consequently, alternative models of a continually cold, arid early Mars, with ice, brines, low atmospheric pressure and groundwater-driven hydrological systems have also been proposed^{20,21}.

Here we review diverse data sets on the composition of Mars obtained over the past decade, focusing on observational constraints on the processes leading to clay mineral formation. Within a conceptual framework for chemical changes and mineral assemblages expected under different alteration scenarios, we examine mineral associations in clay-bearing terrains and review the relative ages and stratigraphic relationships of clay-bearing deposits. Collectively, these data sets show that much aqueous alteration on ancient Mars occurred in the subsurface, with clay production, at low volume ratios of water to rock, by interaction with heated ground waters. Frozen, arid conditions may have been hallmarks of the surface environment since the early-Noachian period, and in this model the proposed warmer and wetter early Mars was largely beneath the surface.

¹Institut d'Astrophysique Spatiale, Université de Paris-Sud XI, 91405 Orsay Cedex, France. ²Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA. ³The Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland 20723, USA. ⁴HydrASA, Université de Poitiers, 86022 Poitiers Cedex, France. ⁵Department of Earth & Planetary Sciences, Washington University in St Louis, St Louis, Missouri 63130, USA. [†]Present address: Division of Geological & Planetary Sciences, California Institute of Technology and Jet Propulsion Laboratory, Pasadena, California 91125, USA.

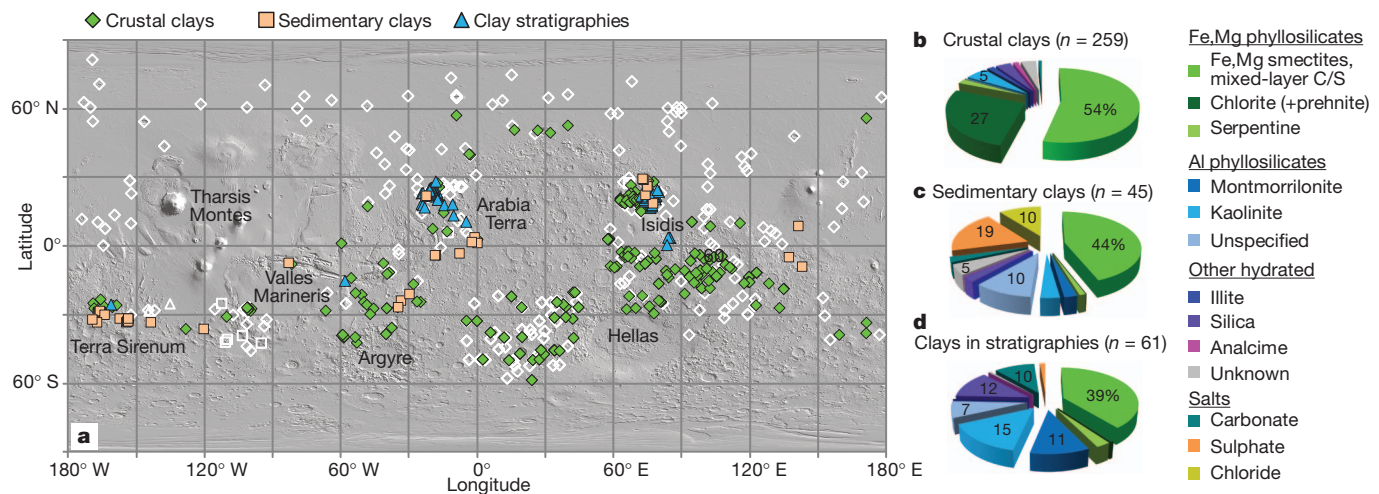


Figure 1 | Clay mineral distribution and diversity on Mars. **a**, CRISM-targeted images surveyed for the presence of clay minerals, grouped by geological setting and superimposed on a shaded relief map. Open symbols mark sites where no clays were found. **b, c, d**, Percentage frequency of detection of alteration phase(s), grouped by geological setting; *n* is the total number of

Controls on clay mineralogy

Waters interacting with rocks alter thermodynamically unstable primary materials such as glasses, olivine, pyroxenes and feldspars, and produce new stable minerals such as clays, salts and oxides (Table 1). Mafic and ultramafic iron and magnesium-rich compositions are typical of Martian rocks of all ages⁹. Given similar starting compositions over most of Mars's surface, thermodynamic variables (for example temperature and pressure), kinetic parameters for mineral formation and destruction, and water chemistry (for example pH, oxidation state and ion activities) control which alteration minerals are produced. Two environmental parameters determining water chemistry are key: the degree to which alteration occurs in systems chemically 'open' or 'closed' to other reservoirs, for example the atmosphere; and the amount of water available to react with the rock, expressed as the water-to-rock ratio (W/R). These parameters control the mobility of elements during alteration, and distinctive mineral assemblages are produced, depending on the alteration environment (Fig. 2).

In closed systems in the subsurface, waters are isolated from the atmosphere and approach chemical equilibrium with precursor-rock chemistry. Temperature and pressure control the resulting alteration mineral assemblages^{22–25}. The water-to-rock ratio is typically low (W/R < 1), and redox, pH and ion activities are determined on the local scale

images within which clay minerals were detected (a total of 365 of 639 images included in this meta-analysis). The percentage given is not areal coverage but rather the number of detections of a given phase divided by the number of detections of all alteration phases within the geological setting.

within the rock pore spaces^{22,26}. Reactions with mafic and ultramafic lithologies control solution chemistry, buffering fluids to be anoxic and alkaline. Transport of soluble ions is limited, so alteration is largely isochemical; that is, although bulk mineralogy changes, elemental abundances remain approximately constant. This is observed for some basalt-hosted sea-floor hydrothermal systems²⁷ and terrestrial hydrothermal systems within Icelandic basalts^{26,28} (Fig. 2). Because elemental abundances of altered rock differ little from the parent rock, mineralogy is key to deciphering the record of aqueous processes. Products of alteration of the bulk rock in systems with negligible sulphur include iron oxides and iron(II)/magnesium smectites, chlorite and serpentine at somewhat higher temperatures, and amphiboles at still higher temperatures^{22–24,26} (>400 °C). Hydrous silicates such as silica, zeolite and prehnite silica precipitate in fractures and pore spaces^{22,23,26}.

By contrast, in open systems nearer the surface, fluids are in communication with an atmospheric reservoir, and atmospheric chemistry exerts strong control on pH, redox and ionic activities of the fluids. At high water-to-rock ratios (W/R > 1) in an open system, transport of elements by leaching changes the bulk elemental composition^{22,25,29–32}. Oxidation and hydrolysis of iron(II) drive waters in contact with Mars's atmosphere to acidic pH³³. For oxic fluids interacting with basalt at moderate pH, silica, chlorine, sodium and potassium are highly soluble

Table 1 | Alteration minerals reported on Mars as of July 2011

Class	Group, mineral or phase	Formula
Phyllosilicates (clay minerals)	Fe,Mg smectites (for example nontronite and saponite)	(Ca, Na) _{0.3–0.5} (Fe,Mg,Al) _{2–3} (Al,Si) ₄ O ₁₀ (OH) ₂
	Aluminium smectite (for example montmorillonite)	(Na,Ca) _{0.33} (Al,Mg) ₂ (Si ₄ O ₁₀)(OH) ₂
	Kaolin group minerals (for example kaolinite and halloysite)	Al ₂ Si ₂ O ₅ (OH) ₄
	Chlorite	(Mg,Fe ²⁺) ₅ Al(Si ₃ Al) ₄ O ₁₀ (OH) ₈
	Serpentine	(Mg,Fe) ₃ Si ₂ O ₅ (OH) ₄
	High-charge Al,K phyllosilicate (for example illite/muscovite)	(K,H ₃ O)(Al,Mg,Fe) ₂ Al _x Si _{4–x} O ₁₀ (OH) ₂
Other hydrated silicates	Prehnite	Ca ₂ Al(AlSi ₃ O ₁₀)(OH) ₂
	Analcime	NaAlSi ₂ O ₆ ·H ₂ O
	Opaline silica	SiO ₂ ·H ₂ O
		(Mg,Fe,Ca)CO ₃
Carbonates	Magnesium, calcium and iron carbonates	(Fe,Mg)SO ₄ ·nH ₂ O
		CaSO ₄ ·2H ₂ O
Sulphates	Fe,Mg mono- and polyhydrated sulphates	KAl ₃ (SO ₄) ₂ (OH) ₆
	Gypsum	KFe(III) ₃ (OH) ₆ (SO ₄) ₂
	Alunite	Fe(III)SO ₄ (OH)
	Jarosite	For example NaCl and MgCl ₂
		(Mg,Ca)(ClO ₄) ₂
Chlorides	Chlorides	Fe ₂ O ₃
Perchlorates	Perchlorates	FeO(OH)
Iron oxides	Haematite	
	Goethite	

Near-infrared vibrational absorptions related to OH, water and CO₂ and electronic transitions related to iron permit the remote identification and mapping of numerous phases. All mineral phases are observed from orbit with imaging spectroscopy, except for chlorides and perchlorates, which have been reported by other orbiting and surface missions. The estimated abundance threshold for detection is ~10%.

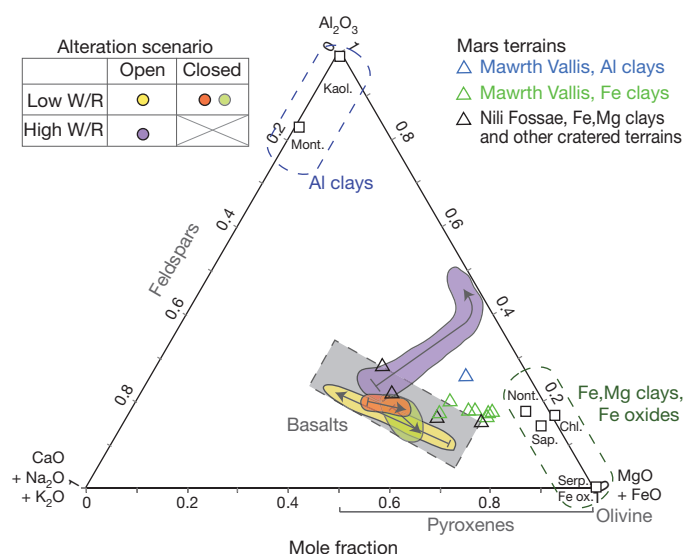


Figure 2 | Chemical and mineralogical changes observed during aqueous alteration of basalt. Rocks of similar composition altered under different conditions show distinctive chemical changes characteristic of the alteration environment. Key parameters are the water-to-rock ratio (W/R) and whether the system is chemically open or closed to the atmosphere. Compositions of altered terrestrial basalts are plotted (after ref. 31) on a ternary diagram with major cations at the apices. The coloured regions encompass individual bulk-rock data points from the references (not shown), and arrows indicate the alteration trajectory. High-W/R, open-system weathering (purple²⁹) leads to substantial chemical fractionation. Alkali and magnesium cations are removed, and primary minerals in rocks are replaced with aluminium clays and iron oxides. Cations liberated and transported during clay-forming reactions may subsequently form salts³². By contrast, low-W/R alteration results in little chemical fractionation and transport, although sample mineralogy changes. In open, low-W/R systems (yellow³¹), directionality of change is governed by olivine dissolution, and coatings of salts and amorphous products are generated by thin-film reactions before water evaporation. In closed, low-W/R systems (orange²⁷ and green²⁸), primary minerals are replaced with Fe,Mg clay minerals and oxides and products such as silica and zeolites may precipitate in pore spaces²⁶. Open triangles represent the chemical composition of clay-bearing units on Mars, calculated from estimates of dust-free modal mineralogy⁷⁴. With the exception of aluminium-clay-bearing units in Mawrth Vallis, these lie on the mixing line between basalt and Fe,Mg clays expected for low-W/R, closed-system-type alteration. Chl., chlorite; Kaol, kaolinite; Mont., montmorillonite; Nont., nontronite; Sap., saponite; Serp., serpentine.

and, hence, mobile. Magnesium, calcium and sometimes silica are relatively mobile, whereas aluminium, iron(III) and titanium are conserved, that is, immobile^{29,30}. Consequently, as weathering proceeds, the bulk chemistry of the residual rock becomes more aluminium rich, and leaching alters mafic lithologies to aluminium clay (for example kaolinite and montmorillonite) and iron oxide assemblages^{22,29} (Fig. 2). From the transported weathering fluids, salts (chlorides, carbonates or sulphates) precipitate. Carbon dioxide or sulphurous gases present in the atmosphere help maintain fluid acidity and also dictate the composition of the salts¹². Oxic surface conditions and/or oxidized sulphurous species generating acidity inhibit Fe,Mg smectite formation, promoting formation of jarosite, silica, iron oxides, amorphous products^{12,34,35} and, if more water is available, kaolinite³⁶. If less water is available (W/R < 1), the most soluble primary phase, olivine, dissolves first, moderately depleting the residual phase in magnesium and iron³¹, but little additional fractionation or alteration of the parent rock occurs. Transport to segregate fluids cannot occur in spite of the open nature of the system, and amorphous products and salts form as coatings and rinds from weathering by thin films of water³¹.

Armed with this conceptual framework for prediction of mineral assemblages under different environmental conditions, the composition of clay-bearing units on Mars can be compared with the above scenarios and used to determine the conditions under which alteration occurred.

Clay distribution and diversity

The OMEGA and CRISM visible/near-infrared orbital imaging spectrometers, on board the Mars Express and Mars Reconnaissance Orbiter spacecraft, respectively, have revealed thousands of outcrops bearing clay minerals distributed throughout the southern highlands⁸ and exposed by large craters that penetrated overlying lavas of the northern lowlands³⁷. The distribution of clays on Mars (Fig. 1) is controlled principally by exposure. Hesperian or later lavas and mantling materials typically overlie globally widespread buried clays, except where tectonism, erosion or cratering exposes underlying units. In a meta-analysis of results from examination of 639 CRISM high-resolution images^{11,37–54}, we catalogued the types and distribution of clay minerals and co-occurring alteration phases. Occurrences were classified into one of three geological settings: ‘crustal’, describing ancient terrains resulting from igneous processes and degraded by impact cratering (‘deep phyllosilicates’ in the terminology of ref. 11); ‘sedimentary’, describing fan-shaped or layered deposits within basins with a setting clearly pointing to transport of the materials (‘phyllosilicates in intracrater fans’, ‘plains sediments’, and ‘intracrater clay sulphate deposits’¹¹); and ‘in stratigraphies’, describing multiple units where clays and other alteration minerals are found in contact with each other in a defined stratigraphy but where either a crustal or sedimentary origin cannot be assigned to all of the units (‘layered phyllosilicates’ or ‘carbonate deposits’¹¹) (Fig. 1).

Across all geological settings, Fe,Mg smectites are the most commonly occurring clay mineral. Compositions range from nontronite (iron end-member) to saponite (magnesium endmember)⁸, and may in some cases include physical mixtures of chlorite and smectite or mixed-layer chlorite–smectite clays, which are mineralogically and structurally distinct but spectrally difficult to distinguish⁴⁸. Notably, the other alteration minerals accompanying Fe,Mg smectites vary markedly depending on geological setting (Fig. 1b–d).

Crustal clays

Fe,Mg smectite and/or chlorite are nearly always present in crustal clay units (detected in 78% and 39% of exposures, respectively). Chlorite is sometimes accompanied by the spectrally similar but mineralogically distinct phase prehnite, which forms at temperatures between 200 and 400 °C (refs 37, 39, 41). Distinctive associations of minerals vary from crater to crater, and can include silica, analcime, serpentine and illite/muscovite accompanying smectites and chlorite^{37,41–43}. These minerals and the assemblages in which they are found indicate a range of formation temperatures from ambient to a maximum of 400 °C^{37,41,55}. A few heavily eroded terrains near the Argyre basin³⁹, Terra Sirenum^{54,56}, the Claritas Fossae⁴² and the Nili Fossae^{8,41,57} host clays within geographically extensive Noachian units exposed at the surface, but most crustal clays are detected in the walls, ejecta or central peaks of impact craters⁸.

It is not known whether crustal clays associated with craters are generated as a result of impacts or are pre-existing compositions excavated from depth. Laboratory studies so far have shown that instantaneous shock does not form clays⁵⁸. However, modelling^{59,60} and terrestrial studies⁶¹ demonstrate that heat from an impact can initiate hydrothermal systems. For the sizes of craters with which clays are associated (diameters of ~20–200 km in this study), hydrothermal systems may last thousands of years at depth^{59,60}. These might produce clays near the hot central peak, beneath an impact melt sheet or in the lower crater walls⁶². For the Martian craters studied so far, the alteration minerals associated with ejecta and the upper walls, where hydrothermal circulation is not predicted, are the same as those associated with the central peak regions^{6,8,37,51,63}. Although morphologies consistent with post-impact hydrothermal activity have been proposed⁶⁴, collectively, observations are consistent with the greater relative importance of the excavation of pre-existing clay-bearing rocks. Materials emplaced ballistically as crater ejecta or constituting the central peaks originate from depths of up to about one-tenth the crater diameter⁶⁵, or up to ~5–10 km for the craters in this study.

Sedimentary clays

Clays and salts have been detected in several of Mars's many putative palaeolakes and fluvial basins. In marked contrast with clays in crustal settings, the hydrated silicates prehnite, analcime, illite/muscovite and chlorite (except possibly as mixed with smectite) have not been detected. Instead, mineral associations vary according to the fluvial system. Open-basin sediments in the Jezero system host Fe,Mg smectite clays and magnesium carbonate⁴¹. Holden⁴⁸, Eberswalde⁴⁸ and other craters⁶⁶ contain predominantly Fe,Mg smectite clays. Broad, intercrater depressions in the southern hemisphere with poor drainage connectivity have distinctive units with chloride salt units overlying Fe,Mg clays^{11,44,67}. Putative lacustrine sediments within the closed basins of Gale, Columbus and Cross craters contain sulphate salts, accompanied by aluminium clays and/or nontronite^{47,54}. From the available data it is not possible to determine definitively whether the clay minerals are allochthonous (transported) or autochthonous (formed *in situ*). Nevertheless, basins fed by valleys typically host minerals closely comparable to those in nearby highland watersheds, consistent with detrital origin^{41,48}. By contrast, salts found in closed basins are not identifiable in the surrounding terrains, and thus instead indicate *in situ* precipitation.

Clays in stratigraphies

A few geographically extensive stratigraphic sections preserve contacts between different clay-bearing rock compositions. By contrast with sedimentary clays found in low-lying depositional basins or their eroded remnants, clays in stratigraphy occur in high-standing topography, and the contacts between clay units typically follow a pre-existing

topographic surface. The nature of this contact suggests formation by either *in situ* alteration, that is, leaching, of pre-existing surface materials, or draping by later air-fall deposits. Hence, these clays in stratigraphy cannot be defined as distinctly crustal or sedimentary. A characteristic feature is the presence of aluminium clays overlying Fe,Mg clays (Figs 1d and 3). In fact, except for a few detections in materials disrupted by cratering⁴² and in sedimentary basins^{54,66}, almost all aluminium clays detected so far on Mars are found in upper, late-Noachian clay-bearing strata above Fe,Mg clays. In Mawrth Vallis and greater Arabia Terra, kaolinite, montmorillonite and silica overlie a nontronite-bearing unit^{6,50,68–70}. Around the Nili Fossae, a thin kaolinite-bearing unit overlies both crustal and sedimentary Fe,Mg smectite clay units. In the eastern portion of this region, magnesium carbonate and serpentine associated with an olivine-rich unit occupy a stratigraphic position similar to that of kaolinite^{41,42,57}. Aluminium clays overlying Fe,Mg clays are also exposed high in the walls of Valles Marineris¹¹ (Fig. 3) and in other scattered exposures across the southern highlands⁵⁶.

Establishing timing

For six regions of Mars^{11,39,46,47,49,50,53,68–70}, the compositional stratigraphy of clay-bearing units has been defined relative to regional geological units for which surface age has been determined from crater counts, making it possible to correlate stratigraphic sections from different parts of the planet (Fig. 3). The deepest portions of the stratigraphies include discrete units containing Fe,Mg smectites, chlorites or their mixtures and are a few hundred metres to kilometres in thickness. Their morphology ranges from discontinuous, quasi-horizontal layers, found at

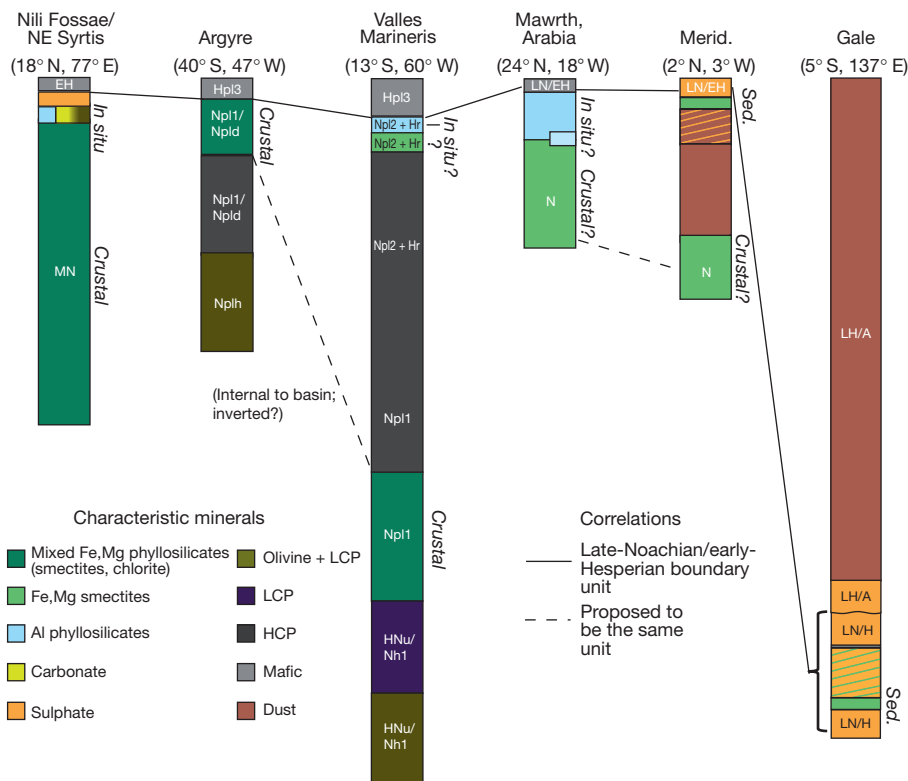


Figure 3 | Compositional stratigraphy of clay-bearing units. Published regional stratigraphies from different locations on the planet—Nili Fossae/northeastern Syrtis^{41,49,57}, Argyre³⁹, Valles Marineris¹¹, Mawrth Vallis and Arabia Terra^{46,50,68–70}, Terra Meridiani⁵³, and sedimentary clays at Gale crater⁴⁷ (latitudes and longitudes as shown)—have been compiled and correlated. Columns, except those for Valles Marineris and Gale, are drawn on the same vertical scale. Section thickness ranges from ~250 m, at Mawrth Vallis, to 8 km, at Valles Marineris. Characteristic minerals or compositions that characterize each unit in near-infrared spectral data are shown. Hatching indicates the interbedding of two minerals in the same large-scale unit, and colour gradation indicates that the corresponding minerals occupy the same unit, grading from

one to the other at various points. Upper units have been dated to the Noachian/Hesperian boundary and can be correlated globally. Correlations between lower stratigraphic units can sometimes be inferred on a regional basis^{39,50}. Unit labels correspond to inferred age (EH, early Hesperian; LN, late Noachian; MN, middle Noachian; A, Amazonian) or assigned global geological map units (for example Np1) identified in the section by the authors of the original published stratigraphy, where the first letter indicates the age of the rock unit. The geological setting of clay formation is given in italics, and, where consensus does not exist in the literature, a question mark is appended. '*In situ*' means in-place formation of the stratigraphic unit by *in situ* alteration; sed., sedimentary. HCP, high-calcium pyroxene; LCP, low-calcium pyroxene.

Mawrth Vallis⁶⁹, to brecciated deposits with heterogeneous 10^0 – 10^3 -m-scale blocks bearing Fe,Mg smectite or low-calcium pyroxene in crust disrupted by Isidis basin formation⁵⁷. Volumetrically large units with spectral signatures of olivine and low- and high-calcium pyroxene but no signs of hydrous minerals are also found in the deepest sections exposed around the Argyre basin and in Valles Marineris^{11,39}. As was previously found using OMEGA data⁷, stratigraphic relations in CRISM data show that nearly all sulphates occur in units younger than the Fe,Mg clays. CRISM data also resolve a similar relationship for multiple clays in stratigraphic section: relatively thin (<50-m) aluminium-clay-bearing units lie above Fe,Mg smectites and chlorites. They are constrained in timing by crater count dating of overlying unaltered late-Noachian/early-Hesperian mafic rocks above and Noachian units beneath. No Fe,Mg clays stratigraphically above aluminium clays or with interbedding of aluminium and Fe,Mg clay units has been found.

We next consider how sedimentary clays fit into the temporal-stratigraphic framework. The lower strata of the Gale crater mound are sedimentary and were probably emplaced during the late Noachian⁴⁷ (Fig. 3). Crater count data show that clay- and sulphate-bearing sediments were emplaced in Cross and Columbus craters in the late Noachian/early Hesperian⁵⁴. Similarly, most occurrences of smectite chloride deposits correspond to this time period^{44,67}. The ages of these sedimentary clay deposits agree well with ages assigned previously to geomorphic features indicative of surface waters. Most valley networks were last active at the late-Noachian/early-Hesperian boundary¹⁶, when

fluvial processes intensified relative to the early-to-middle-Noachian period^{71,72}. Thus, stratigraphy and crater counts together show that the timing of sedimentary clay deposition, surface fluvial activity and the period of aluminium clay formation are coincident within present age-discrimination capabilities and post-date formation of many of the thicker, deep Fe,Mg clay units (Fig. 4).

Early Martian environments

Considered in the context of scenarios for aqueous alteration, evidence from the Martian rock record (Fig. 2) indicates that most clay minerals—specifically, those comprising Fe,Mg clay mineral units deep in the crust—formed in the subsurface in closed systems at temperatures ranging from ambient to low-grade hydrothermal (<400 °C). This formation mechanism is indicated by five lines of evidence: (i) observed mineral assemblages of Fe,Mg smectites and chlorite, which form under anoxic, alkaline, high-pH conditions^{22,24,34,35}; (ii) the presence of the accompanying phases prehnite, analcime, serpentine and illite/muscovite, which form at higher temperatures, ranging up to 400 °C (refs 22, 23, 37, 39, 41, 55); (iii) the pervasive presence of clay minerals in large volumes of Mars's deepest exposed crustal materials in spite of (iv) thermal infrared and elemental data consistent with the dominance of primary minerals by weight fraction^{9,44,73,74} (Fig. 2); and (v) the morphology of Noachian fluvial valleys indicating hyperarid surface conditions except during the late Noachian^{71,72}, when episodic floods incised numerous highland valley networks⁷⁵.

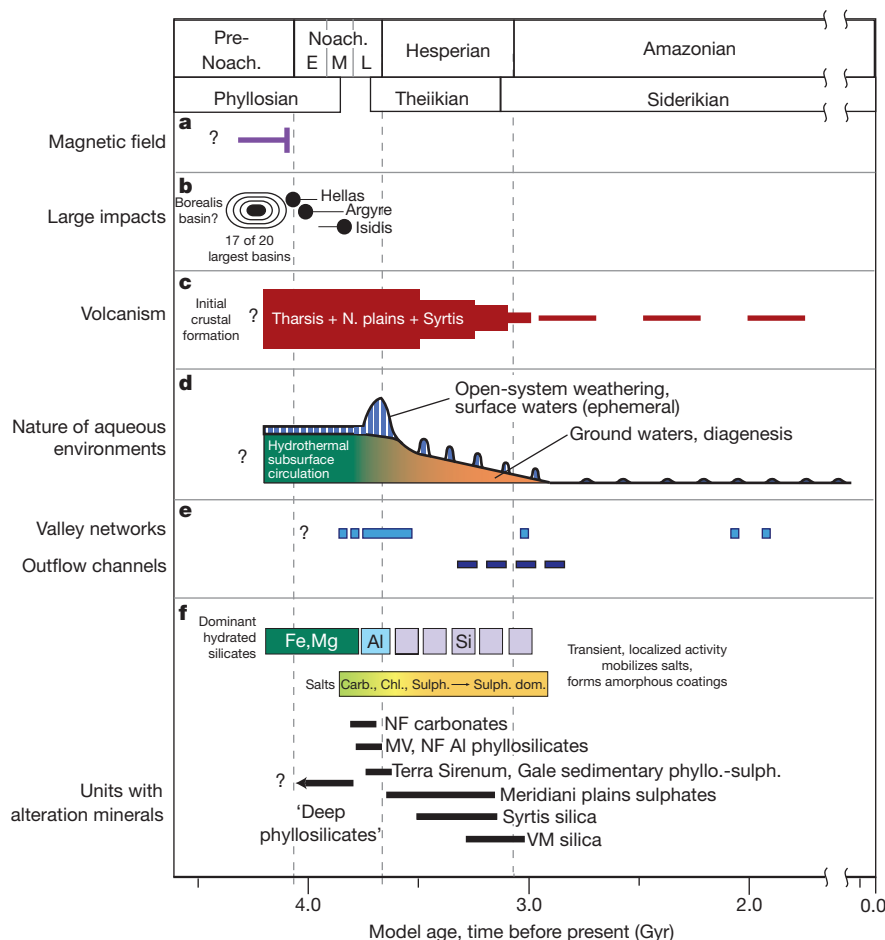


Figure 4 | Timeline of major processes in Mars history. a–c, Major geological processes influencing water availability, including the presence of a magnetic field⁹⁵ (a), impact cratering^{96,97} (b) and volcanism^{86,98} (c). d, Schematic depicting the changing nature of environments hosting liquid water, as implied by the geological evidence discussed herein. e, f, Evidence of liquid water: timing of valley network and outflow channel activity^{15,16} ages of key minerals

formed by aqueous alteration (e) and important regional units with alteration minerals (f). Relative timing is determined using relative crater densities and stratigraphic relationships. Absolute ages of period boundaries⁵ have uncertainties of several hundred million years, inherent to extrapolation from cratering statistics⁹⁹. NF, Nili Fossae; MV, Mawrth Vallis; VM, Valles Marineris; Carb., carbonates; Chl., chlorites; Sulph., sulphates.

The physiochemical conditions for globally widespread Fe,Mg clay formation imply the existence of widespread low-grade metamorphism/diagenesis²³, hydrothermal systems^{25,26} or direct precipitation during cooling of volatile-rich basaltic lavas⁷⁶. For the final mechanism, substantial volatile content in lavas could be obtained by assimilation of ice or liquid water during magma ascent. For the first two mechanisms, subsurface liquid waters, specifically large-scale groundwater aquifers, are required. A long-standing model is that of a Martian cryosphere and melting of subsurface ice at depth⁷⁷. Several mechanisms existed on ancient Mars to melt ice and promote hydrothermal circulation of ground water: large impacts⁶¹, especially during early, heavy impact bombardment⁷⁸; widespread extrusive and intrusive volcanism^{79,80}; and a geothermal gradient⁸¹ greater than that at present.

Implications of widespread Noachian subsurface aqueous activity in thick sections of crust include enhanced rates of crustal cooling⁸¹, possibly significant sequestration of volatiles (for example water and carbon dioxide) in crustal alteration minerals²⁵, and, most importantly, relaxation of the requirements for atmospheric partial pressures adequate to maintain continually warm surface conditions that support long-term liquid water and clay formation by near-surface weathering. Proposed hydrological models driven primarily by ground water rather than precipitation^{20,82} are more consistent with the composition of most Noachian units, which do not show evidence for globally widespread open-system weathering. In the subsurface, and largely out of contact with the atmosphere, low-W/R formation of clays took place, occasionally at increased temperatures. Late-Noachian open-basin lakes^{16,71}, valleys^{71,72,75} and out-flow channels¹⁵ demonstrate at least the episodic presence of liquid water on the Martian surface. Although fluvial degradation was widespread, it implies only local reworking⁷⁵. For most of the Noachian period, surface water derived from snow or ice melt may have quickly entered the subsurface through basaltic rocks made highly permeable on the macroscopic scale by impact- and cooling-induced fractures. This particular model for an early Mars with a warm, wet subsurface and an icy, arid surface simultaneously resolves two paradoxes: the scarcity of salts coeval with clay deposits, in which salts serve as the sink for cations liberated during weathering³², and juvenile-clay mineralogy, that is, the persistence of smectite clays, even at great depths rather than their total diagenetic conversion to illite or chlorite phases⁸³.

On ancient Mars, typical surface conditions may have been dominated by water-ice/water-vapour interactions, much as in the present day, but punctuated by episodic melting to supply liquid water. In the thick Noachian crustal units, there is indeed evidence for sustained diagenesis, the burial-induced transformation of smectite clay minerals, or hydrothermal activity throughout the Noachian, extending into the Hesperian (Fig. 5). Chlorite (which forms directly from hydrothermal fluids or from an Fe(II),Mg smectite precursor) and, more rarely, illite/muscovite (which forms directly from hydrothermal fluids or from an aluminium-rich precursor, given a sufficient supply of potassium) are both found in crustal clays^{23,41}. Abundant salts would not be expected in subsurface, low-W/R alteration because ions are not removed by transport. The fluids are isolated from atmospherically derived volatiles (for example HCO_3^- and SO_4^{2-}), and saturation with respect to secondary aluminosilicates and oxides/hydroxides is favoured over the formation of abundant carbonates, sulphates and chlorides.

In late-Noachian/early-Hesperian boundary units, there is evidence for a period of open-system-style alteration by surface water that may have been intense regionally. Water carved valley networks, eroded and transported highland clays, and redeposited these in sedimentary basins (Fig. 5). The incised, better-integrated, late-Noachian valley networks require higher run-off rates and lower sediment transport, conditions that are reproducible by flow of water over duricrust or partly frozen ground^{72,75}. Clays in stratigraphy may have formed by leaching of surface material to form aluminium clays; after evaporation of water, sulphates and chloride salts precipitated. The salts are found in basins within the southern highlands and—where cation composition can be inferred—seem to be mostly magnesium rich^{11,54,84} as expected given relative cation mobilities^{29,30}.

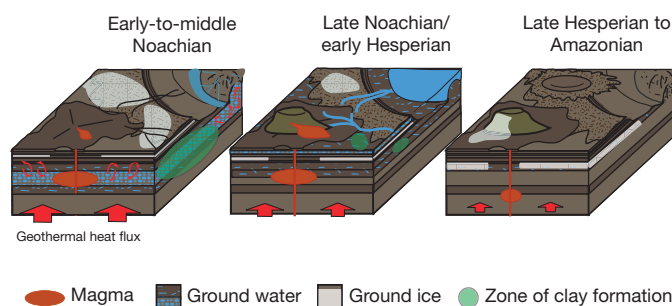


Figure 5 | Evolution of aqueous environments during the first billion years of Mars history. Locations of clay formation are indicated in green. Cold, arid and icy conditions characterized Mars during most of the Noachian, with clay formation mostly in a warmer, wetter subsurface environment. During the late Noachian/early Hesperian, volcanism was widespread and surface waters intermittently carved valleys, sustained lakes, transported sediments to basins and sustained near-surface weathering to form clays near the surface. Conditions from the late Hesperian to the present day have been cold and dry, resulting in the cessation of clay formation.

Seasonal changes coupled with orbital and obliquity parameters may have been sufficient to melt snow and ice and permit surface liquid waters episodically, especially if waters were briny²¹ and atmospheric pressures were slightly higher than at present. Even the densest valley networks require only arid conditions; flows, once initiated, are stable against freezing for many hundreds of kilometres⁸². The key controls on aqueous environments early in Mars's history, permitting a geologically brief period with clement surface conditions at the late-Noachian/early-Hesperian boundary, seem to be heat supplied from volcanism, impacts, crustal cooling and, perhaps, effects of the first two on atmospheric composition (Fig. 4). Temporary release of volatiles by large impacts⁸⁵ or during volcanism^{86,87} might have perturbed the climate to permit liquid surface water over a brief period. Indeed, episodic greenhouse conditions driven by volcanically degassed H_2S and SO_2 (ref. 87) from Tharsis construction and Hesperian plains volcanism might explain the sulphur-rich salts within some closed-basin lakes dating from that period^{47,54}.

As heat-supplying processes declined after Mars's first billion years, so too did the availability of liquid water. When clay formation on a global scale ceased by the Hesperian period, some Al and Fe,Mg smectite clay units were buried to depths of hundreds of metres or more, but little further diagenesis occurred, probably (as previously proposed⁸³) because the availability of ground water decreased in most areas of the planet. By the late Hesperian period, there was neither sufficient surface water for open-system leaching to form clay and salt assemblages nor sufficient subsurface liquid water for hydrothermal circulation or diagenesis, even in the vicinity of the Tharsis volcanoes where evidence of the last sulphate and silica deposits are found¹¹. Major Hesperian units including lavas of the Syrtis Major formation and northern plains are not altered, indicating either that clay formation had ceased by this time or that subsurface alteration occurred at depths $>5\text{--}10\text{ km}$, which are inaccessible to exposure by later cratering or tectonics.

Continuing exploration

The Mars Exploration Rover Opportunity has, as of this writing, arrived at the walls and rim of Endeavour crater, where Fe,Mg clay minerals have been detected using orbital data⁸⁸; and the Mars Science Laboratory rover Curiosity, scheduled for launch in late 2011, will investigate sedimentary clays within Gale crater^{47,89}. The hypothesis that early environments with long-term liquid water were hosted in Mars's subsurface can be tested with data from future missions, most directly by *in situ* investigation of well-defined stratigraphies that include crustal clays. Predictions for our proposed environmental model include (1) direct detection of hydrothermal/metamorphic minerals or mixed-layer clays formed during diagenesis; (2) clay-bearing crustal units with elemental

abundances little changed from that of basalt, implying a low W/R ratio during alteration; (3) textures, similar to those observed in Mars meteorites⁹⁰, with zonation of alteration phases (clays and salts) in veins and vesicles that indicate changing fluid compositions in intrarock flow; and (4) mildly hydrothermal mineral formation temperatures as determined from hydrogen, carbon and oxygen isotopic geothermometers. Additional, indirect evidence supporting an almost continually cold and dry Martian surface could come from measured isotopic ratios and calculation of loss rates of atmospheric gases made possible with future orbital data.

The stratigraphic record shows that most crustal clay units formed before the late Noachian and that all formed earlier than the Hesperian, but it is not known how old the oldest clay mineral occurrences are. Even though Mars's ancient record is better preserved than Earth's, it is still disrupted. Mars's earliest period is typically said to be the Noachian, but no surfaces datable by crater counting survive from before emplacement of the Hellas basin rim⁵. This does not mean that the record of the 'pre-Noachian' is completely lost. Rather, this record exists in the lowermost strata around large impact basins and in megabreccia blocks⁵⁷. Materials comprising these deep units, some of which are crustal clays, have been heavily churned. Nevertheless, the scale of the sometimes layered strata preserved in these 10²–10³-m-scale blocks is comparable to the tens of metres of section explored during the Mars Exploration Rover traverses. Hints of Ca,Fe carbonates mixed with Fe,Mg clays in such blocks, excavated from depth^{45,91}, could indicate earlier, near-surface alteration by waters in contact with an atmosphere. Alternatively, petrographic textures with veins of clays cross-cutting both breccia blocks and their matrix would indicate post-impact hydrothermal systems set up by early, large impacts. The pre-Noachian record of Mars is just beginning to be understood and will require careful examination of small-scale rock units and reconstruction of stratigraphic and temporal context.

Cumulatively, all data acquired so far support the idea that throughout the Noachian and the Hesperian, a persistent and volumetrically extensive aqueous environment existed in the Martian crust. Water activities and chemical gradients may have provided habitable conditions and energy for microbial life in these ancient Martian subsurface hydrothermal systems. Alternatively, intermittent surface waters provided by melting of ice and snow on an arid Martian surface may have sustained a habitable environment. The significant fraction of Earth's microbial biomass within its crust⁹², the thermophilic, chemoautotrophic nature of the last common ancestor of modern life⁹³ and the deep rooting within the phylogenetic tree of organisms from hydrothermal systems⁹⁴ point to the importance of subsurface water–rock reactions in the development of early terrestrial life. On Earth, rocks from the time of life's origins have been destroyed. On Mars, the oldest known aqueous environments in the crust, possibly the oldest habitats, remain preserved for exploration.

- Guggenheim, S. & Martin, R. T. Definition of clay and clay mineral: joint report of the AIPEA nomenclature and CMS Nomenclature Committees. *Clays Clay Miner.* **43**, 255–256 (1995).
- Eberl, D. D., Farmer, V. C. & Barrer, R. M. Clay mineral formation and transformation in rocks and soils. *Phil. Trans. R. Soc. Lond. A* **311**, 241–257 (1984).
- Merriman, R. J. Clay minerals and sedimentary basin history. *Eur. J. Mineral.* **17**, 7–20 (2005).
- Kump, L. R., Brantley, S. L. & Arther, M. A. Chemical weathering, atmospheric CO₂, and climate. *Annu. Rev. Earth Planet. Sci.* **28**, 611–667 (2000).
- Nimmo, F. & Tanaka, K. Early crustal evolution of Mars. *Annu. Rev. Earth Planet. Sci.* **33**, 133–161 (2005).
- Poulet, F. *et al.* Phyllosilicates on Mars and implications for early Martian climate. *Nature* **438**, 623–627 (2005).
This paper details the mineralogy and geological settings of the first clays found on Mars and reports that clay minerals are restricted to Noachian terrains.
- Bibring, J. P. *et al.* Global mineralogical and aqueous Mars history derived from OMEGA/Mars express data. *Science* **312**, 400–404 (2006).
This paper advances the hypothesis that distinctive types of alteration minerals define three sequential environmental epochs in Mars's history.
- Mustard, J. F. *et al.* Hydrated silicate minerals on Mars observed by the Mars reconnaissance orbiter CRISM instrument. *Nature* **454**, 305–309 (2008).
This paper reports that clays on Mars are of diverse mineralogy and geological setting as well as more widespread than previously believed.

- Rogers, D. & Christensen, P. R. Surface mineralogy of Martian low-albedo regions from MGS-TES data: implications for upper crustal evolution and surface alteration. *J. Geophys. Res.* **112**, E01003 (2007).
- Kraft, M. D., Michalski, J. R. & Sharp, T. G. Effects of pure silica coatings on thermal emission spectra of basaltic rocks: considerations for Martian surface mineralogy. *Geophys. Res. Lett.* **30**, 2288 (2003).
- Murchie, S. L. *et al.* A synthesis of Martian aqueous mineralogy after 1 Mars year of observations from the Mars Reconnaissance Orbiter. *J. Geophys. Res.* **114**, E00D06 (2009).
In this paper, nearly a dozen distinctive chemical environments are reported to be preserved in the rock record of ancient Mars, identified and time-ordered by combining geomorphic data with mineralogical data.
- Chevrier, V., Poulet, F. & Bibring, J.-P. Early geochemical environment of Mars as determined from thermodynamics of phyllosilicates. *Nature* **448**, 60–63 (2007).
In this paper, on the basis of thermodynamic arguments a hypothesis is put forward that explains the transition from clay-forming conditions to sulphate-forming conditions in terms of loss of non-carbon-dioxide atmospheric greenhouse gases.
- Andrews-Hanna, J. *et al.* Early Mars hydrology: Meridiani playa deposits and the sedimentary record of Arabia Terra. *J. Geophys. Res.* **115**, E06002 (2010).
- Haberle, R. M. *et al.* On the possibility of liquid water on present-day Mars. *J. Geophys. Res.* **106**, E10 (2001).
- Tanaka, K. L. Sedimentary history and mass flow structures of Chryse and Acidalia Planitiae, Mars. *J. Geophys. Res.* **102**, 4131–4149 (1997).
- Fassett, C. I. & Head, J. W. The timing of Martian valley network activity: constraints from buffered crater counting. *Icarus* **195**, 61–89 (2008).
- Haberle, R. Early Mars climate models. *J. Geophys. Res.* **103**, E12 (1998).
- Halevy, I. *et al.* A sulfur dioxide climate feedback on early Mars. *Science* **318**, 1903–1907 (2007).
- Phillips, R. *et al.* Massive CO₂ ice deposits sequestered in the south polar layered deposits of Mars. *Science* **332**, 838–841 (2011).
- Squyres, S. W. & Kasting, J. F. Early Mars: how warm and how wet? *Science* **265**, 744–749 (1994).
- Fairén, A. G., Davila, A. F., Gag-Duport, L., Amils, R. & McKay, C. P. Stability against freezing of aqueous solutions on early Mars. *Nature* **459**, 401–404 (2009).
- Meunier, A. *Clays* (Springer, 2005).
- Frey, M. & Robinson, D. *Low-Grade Metamorphism* (Blackwell, 1999).
- Spear, F. S. *Metamorphic Phase Equilibria and Pressure-Temperature-Time Paths* (Mineralogical Society of America, 1993).
- Griffith, L. L. & Shock, E. L. Hydrothermal hydration of Martian crust: illustration via geochemical model calculations. *J. Geophys. Res.* **102**, 9135–9143 (1997).
- Franzson, H., Zierenberg, R. & Schiffman, P. Chemical transport in geothermal systems in Iceland: evidence from hydrothermal alteration. *J. Volcanol. Geotherm. Res.* **173**, 217–229 (2008).
- Cann, J. R. & Vine, F. J. An area on the crest of the Carlsberg Ridge: petrology and magnetic survey. *Phil. Trans. R. Soc. Lond. A* **259**, 198–217 (1966).
- Ehlmann, B. L., Mustard, J. F. & Bish, D. L. in *Analogue Sites for Mars Missions: MSL and Beyond* abstr. 6020, (<http://www.lpi.usra.edu/meetings/analogue2011/pdf/6020.pdf>) (Lunar and Planetary Institute, 2011).
- Nesbitt, H. W. & Wilson, R. E. Recent chemical weathering of basalts. *Am. J. Sci.* **292**, 740–777 (1992).
- Gislason, S. R., Arnorsson, S. & Armannsson, H. Chemical weathering of basalt in southwest Iceland: effects of runoff, age of rocks, and vegetative/glacial cover. *Am. J. Sci.* **296**, 837–907 (1996).
- Hurowitz, J. A. & McLennan, S. L. A ~3.5 Ga record of water-limited, acidic weathering conditions on Mars. *Earth Planet. Sci. Lett.* **260**, 432–443 (2007).
This paper uses a conceptual framework from the Earth science literature to understand element transport in different alteration scenarios and shows acidic, low-W/R alteration since the Hesperian.
- Milliken, R. E. *et al.* Missing salts on early Mars. *Geophys. Res. Lett.* **36**, L11202 (2009).
This paper models Noachian clay formation in open-system weathering and notes that abundant coeval salts should be produced but are not detected.
- Hurowitz, J. *et al.* Origin of acidic surface waters and the evolution of atmospheric chemistry on early Mars. *Nature Geosci.* **3**, 323–326 (2010).
- Harder, H. Nontronite synthesis at low temperatures. *Chem. Geol.* **18**, 169–180 (1976).
- Tosca, N. J. *et al.* in *Workshop on Martian Phyllosilicates: Recorders of Aqueous Processes?* abstr. 7030, (http://www.ias.u-psud.fr/Mars_Phyllosilicates/phyllo/4.Wednesday%20afternoon/6.Tosca_Phyllo_2008.ppt) (Institut d'Astrophysique Spatiale, 2008).
- Altheide, T. *et al.* Mineralogical characterization of acid weathered phyllosilicates with implications for secondary Martian deposits. *Geochim. Cosmochim. Acta* **74**, 6232–6248 (2010).
- Carter, J., Poulet, F., Bibring, J.-P. & Murchie, S. Detection of hydrated silicates in crustal outcrops in the northern plains of Mars. *Science* **328**, 1682–1686 (2010).
- Baldrige, A. M. *et al.* Contemporaneous deposition of phyllosilicates and sulfates: using Australian acidic saline lake deposits to describe geochemical variability on Mars. *Geophys. Res. Lett.* **36**, L19201 (2009).
- Buczkowski, D. L. *et al.* Investigation of an Argyre basin ring structure using Mars Reconnaissance Orbiter/Compact Reconnaissance Imaging Spectrometer for Mars. *J. Geophys. Res.* **115**, E12011 (2010).
- Buczkowski, D. *et al.* in *41st Lunar Planet. Sci. Conf.* abstr. 1458, (<http://www.lpi.usra.edu/meetings/lpsc2010/pdf/1458.pdf>) (Lunar and Planetary Institute, 2010).

41. Ehlmann, B. L. *et al.* Identification of hydrated silicate minerals on Mars using MRO-CRISM: geologic context near Nili Fossae and implications for aqueous alteration. *J. Geophys. Res.* **114**, E00D08 (2009).
- This paper provides the first report of minerals in characteristic assemblages (diagenetic, low-grade metamorphic and hydrothermal) indicative of alteration at increased temperatures, ranging from above ambient to 400 °C.**
42. Ehlmann, B. L., Mustard, J. F. & Murchie, S. L. Geologic setting of serpentine deposits on Mars. *Geophys. Res. Lett.* **37**, L06201 (2010).
43. Fraeman, A. A. *et al.* in *40th Lunar Planet. Sci. Conf.* abstr. 2320, (<http://www.lpi.usra.edu/meetings/lpsc2009/pdf/2320.pdf>) (Lunar and Planetary Institute, 2009).
44. Glotch, T. D. *et al.* Distribution and formation of chlorides and phyllosilicates in Terra Sirenum, Mars. *Geophys. Res. Lett.* **37**, L16202 (2010).
45. Michalski, J. R. & Niles, P. B. Deep crustal carbonate rocks exposed by meteor impact on Mars. *Nature Geosci.* **3**, 751–755 (2010).
46. McKeown, N. *et al.* Characterization of phyllosilicates observed in the central Mawrth Vallis region, Mars, their potential formational processes, and implications for past climate. *J. Geophys. Res.* **114**, E00D10 (2009).
47. Milliken, R. E. *et al.* Paleoclimate of Mars as captured by the stratigraphic record in Gale Crater. *Geophys. Res. Lett.* **37**, L04201 (2010).
48. Milliken, R. E. & Bish, D. L. Sources and sinks of clay minerals on Mars. *Phil. Mag.* **90**, 2293–2308 (2010).
49. Mustard, J. F. & Ehlmann, B. L. in *42nd Lunar Planet. Sci. Conf.* abstr. 2355, (<http://www.lpi.usra.edu/meetings/lpsc2011/pdf/2355.pdf>) (Lunar and Planetary Institute, 2011).
50. Noe Dobra, E. Z. *et al.* Mineralogy and stratigraphy of phyllosilicate-bearing and dark mantling units in the greater Mawrth Vallis/west Arabia Terra area: constraints on geological origin. *J. Geophys. Res.* **115**, E00D19 (2010).
51. Roach, L. H. *et al.* Hydrated mineral stratigraphy of Ius Chasma, Valles Marineris. *Icarus* **206**, 253–268 (2010).
52. Wiseman, S. M. *et al.* Phyllosilicate and sulfate-hematite deposits within Miyamoto crater in southern Sinus Meridiani, Mars. *Geophys. Res. Lett.* **35**, L19204 (2008).
53. Wiseman, S. M. *et al.* Spectral and stratigraphic mapping of hydrated sulfate and phyllosilicate-bearing deposits in northern Sinus Meridiani, Mars. *J. Geophys. Res.* **115**, E00D18 (2010).
54. Wray, J. J. *et al.* Columbus crater and other possible groundwater-fed paleolakes of Terra Sirenum, Mars. *J. Geophys. Res.* **116**, E01001 (2011).
55. Ehlmann, B. L., Mustard, J. F., Clark, R. N., Swayze, G. A. & Murchie, S. L. Evidence for low-grade metamorphism, hydrothermal alteration, and diagnosis on Mars from phyllosilicate mineral assemblages. *Clays Clay Miner.* **59**, 357–375 (2011).
56. Wray, J. J. *et al.* Diverse aqueous environments on ancient Mars revealed in the southern highlands. *Geology* **37**, 1043–1046 (2009).
57. Mustard, J. F. *et al.* Composition, morphology, and stratigraphy of Noachian crust around the Isidis basin. *J. Geophys. Res.* **114**, E00D12 (2009).
58. Allen, C. C., Jercinovic, M. J., See, T. & Keil, K. Experimental shock lithification of water-bearing rock powders. *Geophys. Res. Lett.* **9**, 1013–1016 (1982).
59. Rathbun, J. A. & Squyres, S. W. Hydrothermal systems associated with Martian impact craters. *Icarus* **157**, 362–372 (2002).
60. Abramov, O. & Kring, D. A. Impact-induced hydrothermal activity on early Mars. *J. Geophys. Res.* **110**, E12S09 (2005).
61. Newsom, H. E. Hydrothermal alteration of impact melt sheets with implications for Mars. *Icarus* **44**, 207–216 (1980).
62. Schwenzer, S. P. & Kring, D. A. Impact-generated hydrothermal systems capable of forming phyllosilicates on Noachian Mars. *Geology* **37**, 1091–1094 (2009).
63. Fairén, A. G. *et al.* Noachian and more recent phyllosilicates in impact craters on Mars. *Proc. Natl Acad. Sci. USA* **107**, 12,095–12,100 (2010).
64. Marz, G. A. *et al.* Evidence for Hesperian impact-induced hydrothermalism on Mars. *Icarus* **208**, 667–683 (2010).
65. Melosh, J. *Impact Cratering: A Geologic Process* (Oxford Univ. Press, 1989).
66. Hughes, A. C. G. *et al.* in *42nd Lunar Planet. Sci. Conf.* abstr. 2301, (<http://www.lpi.usra.edu/meetings/lpsc2011/pdf/2301.pdf>) (Lunar and Planetary Institute, 2011).
67. Osterloo, M. M. *et al.* Geologic context of proposed chloride-bearing materials on Mars. *J. Geophys. Res.* **115**, E10012 (2010).
68. Loizeau, D. *Etude Spectrale et Géologique des Phyllosilicates de Mars* 165–209. PhD thesis, Univ. Paris-Sud XI (2008).
69. Michalski, J. *et al.* The Mawrth Vallis region of Mars: a potential landing site for the Mars Science Laboratory (MSL) mission. *Astrobiology* **10**, 687–703 (2010).
70. Bishop, J. L. *et al.* Phyllosilicate diversity and past aqueous activity revealed at Mawrth Vallis, Mars. *Science* **321**, 830–833 (2008).
71. Irwin, R. P. III *et al.* An intense terminal epoch of widespread fluvial activity on early Mars: 2. Increased runoff and paleolake development. *J. Geophys. Res.* **110**, E12S15 (2005).
72. Howard, A. D., Moore, J. M. & Irwin, R. P. An intense terminal epoch of widespread fluvial activity on early Mars: 1. Valley network incision and associated deposits. *J. Geophys. Res.* **110**, E12S14 (2005).
73. Taylor, G. J. *et al.* Mapping Mars geochemically. *Geology* **38**, 183–186 (2010).
74. Poulet, F. *et al.* Abundance of minerals in the phyllosilicate-rich units on Mars. *Astron. Astrophys.* **487**, L41–L44 (2008).
75. Barnhart, C. J. *et al.* Long-term precipitation and late-stage valley network formation: landform simulations of Parana Basin, Mars. *J. Geophys. Res.* **114**, E01003 (2009).
76. Meunier, A. *et al.* The Fe-rich clay microsystems in basalt-komatiite lavas: importance of Fe-smectites for pre-biotic molecule catalysis during the Hadean eon. *Orig. Life Evol. Biosph.* **40**, 253–272 (2010).
77. Clifford, S. M. A model for the hydrologic and climatic behavior of water on Mars. *J. Geophys. Res.* **98**, 10,973–11,016 (1993).
78. Abramov, O. & Mojzsis, S. J. Microbial habitability of the Hadean Earth during the late heavy bombardment. *Nature* **459**, 419–422 (2009).
79. Gulick, V. C. Magmatic intrusions and a hydrothermal origin for fluvial valleys on Mars. *J. Geophys. Res.* **103**, 19365–19387 (1998).
80. Harrison, K. P. & Grimm, R. E. Controls on Martian hydrothermal systems: application to valley network and magnetic anomaly formation. *J. Geophys. Res.* **107**, 5025 (2002).
81. Parmentier, E. M. & Zuber, M. T. Early evolution of Mars with mantle compositional stratification or hydrothermal crustal cooling. *J. Geophys. Res.* **112**, E02007 (2007).
82. Carr, M. H. Stability of streams and lakes on Mars. *Icarus* **56**, 476–495 (1983).
83. Tosca, N. J. & Knoll, A. H. Juvenile chemical sediments and the long term persistence of water at the surface of Mars. *Earth Planet. Sci. Lett.* **286**, 379–386 (2009).
- This paper points out an apparent paradox: if water was readily available and long-lived on Mars, diagenesis/burial metamorphism should have transformed smectites to higher-temperature phases.**
84. Ehlmann, B. L. *et al.* Orbital identification of carbonate-bearing rocks on Mars. *Science* **322**, 1828–1832 (2008).
85. Segura, T. L. *et al.* Environmental effects of large impacts on Mars. *Science* **298**, 1977–1980 (2002).
86. Phillips, R. J. *et al.* Ancient geodynamics and global-scale hydrology on Mars. *Science* **291**, 2587–2591 (2001).
87. Johnson, S. S. *et al.* Sulfur-induced greenhouse warming on early Mars. *J. Geophys. Res.* **113**, E08005 (2008).
88. Wray, J. J. *et al.* Phyllosilicates and sulfates at Endeavour Crater, Meridiani Planum, Mars. *Geophys. Res. Lett.* **36**, L21201 (2009).
89. Grant, J. A. *et al.* The science process for selecting the landing site for the 2011 Mars Science Laboratory. *Planet. Space Sci.* **59**, 1114–1127 (2011).
90. Bridges, J. C. *et al.* Alteration assemblages in Martian meteorites: implications for near-surface processes. *Space Sci. Rev.* **96**, 365–392 (2001).
91. Wray, J. J. *et al.* in *42nd Lunar Planet. Sci. Conf.* abstr. 2635, (<http://www.lpi.usra.edu/meetings/lpsc2011/pdf/2635.pdf>) (Lunar and Planetary Institute, 2011).
92. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: the unseen majority. *Proc. Natl Acad. Sci. USA* **95**, 6578–6583 (1998).
93. Pace, N. R. A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–740 (1997).
94. Reysenbach, A.-L. & Shock, E. Merging genomes with geochemistry in hydrothermal systems. *Science* **296**, 1077–1082 (2002).
95. Acuña, M. H. *et al.* Global distribution of crustal magnetization discovered by the Mars Global Surveyor MAG/ER experiment. *Science* **284**, 790–793 (1999).
96. Frey, H. Ages of very large impact basins on Mars: implications for the late heavy bombardment in the inner solar system. *Geophys. Res. Lett.* **35**, L13203 (2008).
97. Werner, S. C. The early Martian evolution—constraints from basin formation ages. *Icarus* **195**, 45–60 (2008).
98. Werner, S. C. The global Martian volcanic evolutionary history. *Icarus* **201**, 44–68 (2009).
99. Hartmann, W. K. & Neukum, G. Cratering chronology and the evolution of Mars. *Space Sci. Rev.* **96**, 165–194 (2001).

Acknowledgements Thanks to J. Catalano and C. Fassett for detailed feedback on earlier versions of this work; to R. Arvidson, J. Carter, J. Michalski, F. Poulet and M. Vincendon for science discussions; and to the Mars Express and Mars Reconnaissance Orbiter teams for their data collection efforts.

Author Contributions B.L.E. compiled the data sets and led formulation of the manuscript text and figures and the concepts therein. J.F.M. contributed to the text and to the development of hypotheses and scenarios early in manuscript development. S.L.M. contributed to the text and figures and was instrumental in leading data set collection. J.-P.B. and A.M. contributed to formulation of key ideas in the manuscript regarding timing of clay formation relative to other events in Martian history (J.-P.B.) and methods used for assessing clay formation environment (A.M.). A.A.F. and Y.L. helped with text formulation and structure, and A.A.F. provided her database with analyses of hundreds of CRISM images for incorporation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and should be addressed to B.L.E. (ehlmann@caltech.edu).

Motor antagonism exposed by spatial segregation and timing of neurogenesis

Marco Tripodi¹, Anna E. Stepien¹ & Silvia Arber¹

Walking is a key motor behaviour of limbed animals, executed by contraction of functionally antagonistic muscle groups during swing and stance phases. Nevertheless, neuronal circuits regulating the activation of antagonistic extensor–flexor muscles remain poorly understood. Here we use monosynaptically restricted trans-synaptic viruses to elucidate premotor anatomical substrates for extensor–flexor control in mice. We observe a medio–lateral spatial segregation between extensor and flexor premotor interneurons in the dorsal spinal cord. These premotor interneuron populations are derived from common progenitor domains, but segregate by timing of neurogenesis. We find that proprioceptive sensory feedback from the periphery is targeted to medial extensor premotor populations and is required for extensor-specific connectivity profiles during development. Our findings provide evidence for a discriminating anatomical basis of antagonistic circuits at the level of premotor interneurons, and point to synaptic input and developmental ontogeny as key factors in the establishment of circuits regulating motor behavioural dichotomy.

Walking represents a fundamental manifestation of motor behaviour and is based on the selective control of functionally antagonistic muscles. This seemingly simple behaviour has been the subject of scientific debate for more than a century^{1–4}. An important entry point to understand differential regulation of motor output has been the comparative analysis of antagonistic motor neuron pool function. For example, extensor and flexor motor neuron pools in the mammalian spinal cord innervate distinct limb muscles, which are generally active in alternation between an ‘on-ground’ stance and an ‘off-ground’ swing phase during walking⁵. Developmental and genetic studies have addressed the role of neuronal populations derived from transcriptionally defined progenitor domains in motor output regulation, providing insight into the genetic identity of neurons required for left–right but not extensor–flexor regulation^{6–8}.

Functional studies on the regulation of extensor and flexor motor neuron pools provide evidence for selective tuning of antagonistic circuits, most prominently featuring the concept of spinal central pattern generators and sensory feedback loops. Models emerging from this work describe distinct spinal ‘units’ orchestrating stance–swing phase alternation of motor output^{2,7,9–12}. Connection selectivity between functionally antagonistic circuit elements is also evident for sensory input processing to motor neurons and spinal interneurons^{8,13–16}. Group Ia proprioceptive sensory neurons transmit information from peripheral muscle spindles directly to motor neurons and Ia reciprocal interneurons in the ventral spinal cord with high synaptic specificity, obeying rules of functional antagonism^{8,17–21}. Even though these findings illustrate that functional antagonism can be observed at many levels, an overall anatomical assessment of organizational principles of neuronal circuits relaying information to functionally distinct motor neuron pools is currently lacking.

Here we developed a retrograde reconstruction approach of motor output circuitry in the spinal cord to investigate the neuronal ensembles directly connecting to extensor and flexor motor neuron pools, the premotor interneuron populations. We exploited a monosynaptically restricted trans-synaptic rabies virus circuit tracing tool^{22,23} recently developed for its use in the motor output system to study systematically

the distribution of premotor populations²⁴. We provide evidence for segregation of extensor and flexor premotor interneuron populations based on location, time of neurogenesis and synaptic input. These results provide a conceptual framework to understand the regulation of antagonistic motor behaviour at the circuit level.

Spatial segregation of premotor interneurons

To determine the overall spatial organization of spinal premotor interneurons connected to extensor and flexor motor neuron pools (Fig. 1a), we analysed their three-dimensional distribution in the lumbar spinal cord of mice by exploiting trans-synaptic virus-based methods²⁴. To compare the distribution of premotor interneurons connected to motor neurons innervating the ankle extensor muscle gastrocnemius (GS) to the ones controlling the ankle flexor muscle tibialis anterior (TA) (Fig. 1a), we targeted injections of Rab-mCherry to the GS muscle and Rab-eGFP to the TA muscle in postnatal day 3 (p3) mice, each in conjunction with an adeno-associated virus expressing the rabies glycoprotein (AAV-Gly). Premotor interneurons analysed at p12 were found at spinal levels of the source motor neuron pools (lumbar (L) levels L4/5)²⁵, but also many segments rostrally and caudally to it (Fig. 1c, d).

To analyse the three-dimensional distribution, we assigned *x–y–z* coordinates to labelled premotor interneurons (Fig. 1d; thoracic T11 to sacral S1; *n* = 6). Comparison of the overall percentage of labelled interneurons did not reveal significant differences between ipsilateral and contralateral GS and TA premotor interneurons (Fig. 1b; percentage contralateral: GS, 16.3 ± 6; TA, 15.4 ± 5). Dorso-ventral distribution profiles showed the highest density of ipsilateral TA premotor interneurons in a slightly more ventral position than for GS (Fig. 1e), but these differences were not analysed further. We observed the most pronounced divergence in interneuron density along the medio-lateral dimension ipsilateral to injection, with the highest concentration of GS premotor interneurons located significantly more medially than for TA (Figs 1f–h and 2c), a feature independent of the developmental stage of viral injections (Supplementary Fig. 1). To generalize our findings, we analysed the distributions of

¹Biozentrum, Department of Cell Biology, University of Basel, 4056 Basel, and Friedrich Miescher Institute for Biomedical Research, 4058 Basel, Switzerland.

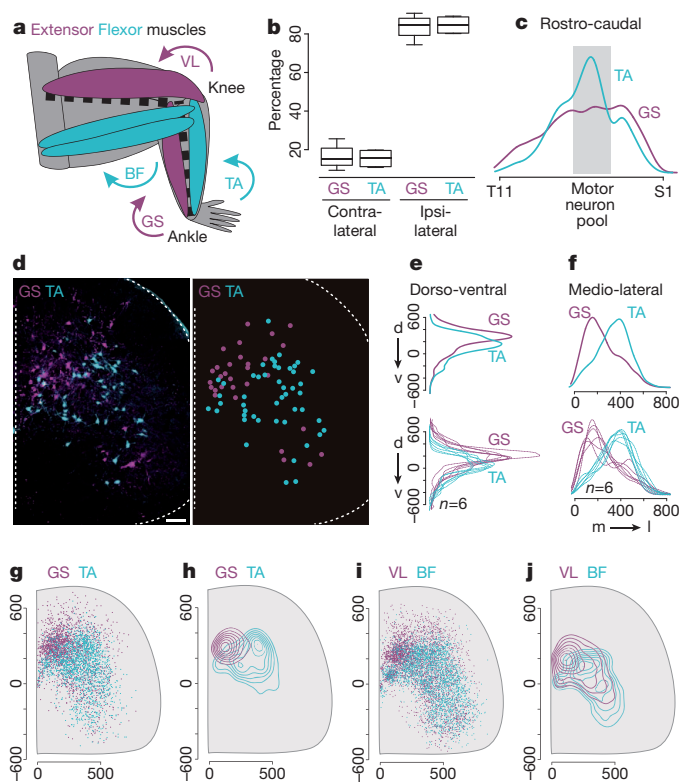


Figure 1 | Spatial segregation of extensor and flexor premotor interneurons. **a**, Scheme of hindlimb ankle and knee extensor (purple) and flexor (turquoise) muscles. BF, posterior biceps femoris; GS, gastrocnemius; TA, tibialis anterior; VL, vastus lateralis. **b**, Percentage of contralateral and ipsilateral GS/TA premotor interneurons. **c**, Rostro-caudal distribution of GS/TA premotor interneuron density (T11–S1). The grey box indicates the motor neuron pool position. **d**, GS and TA premotor interneurons on transverse spinal cord section (left) and digital coordinates (right). Scale bar: 70 μ m. **e**, **f**, Dorso-ventral (**e**) and medio-lateral (**f**) density for GS/TA premotor interneurons. Top: average of $n = 6$ animals, $P_{d-v} < 0.05$, $P_{m-l} < 0.001$, Wilcoxon test; bottom: 6 experiments. **g–j**, Transverse projections of digital reconstructions of individual experiments for GS, TA, VL and BF (**g**, **i**, dots indicate labelled neurons; **h**, **j**, density of neuronal distributions; scale in micrometres).

premotor interneurons connected to two additional pairs of extensor and flexor motor neuron pools: those innervating the knee extensor vastus lateralis (VL) and knee flexor posterior biceps femoris (BF), and those innervating the hip extensor gluteus (GL) and hip flexor iliopsoas (IL) (Fig. 1a; see Methods). Analysis of these additional extensor–flexor premotor interneuron distributions also revealed a clear segregation in the highest density along the medio-lateral axis (Figs 1i, j and 2c and Supplementary Fig. 2). Taken together, these findings indicate that premotor interneuron density segregation along the medio-lateral axis is correlated with the extensor–flexor function rather than with the developmental origin of the innervated muscle (see Methods).

Adductor muscles have a mechanical function distinct from extensor and flexor muscles to move a limb closer to the sagittal body plane, but the adductor muscles of the gracilis (GRA) group are recruited at a point that temporally overlaps with flexor muscle contraction²⁶. Trans-synaptic virus injection into GRA revealed a premotor interneuron distribution pattern similar to flexor distributions (Fig. 2a–c), with the highest density of GRA premotor interneurons laterally and overlapping with the TA premotor interneuron domain (Fig. 2a).

We next performed hierarchical cluster analysis of the distributions of all premotor interneuron ensembles analysed. We found that the flexor premotor interneuron populations TA/BF/IL co-segregate, together with the GRA group, whereas the extensor groups GS/VL/

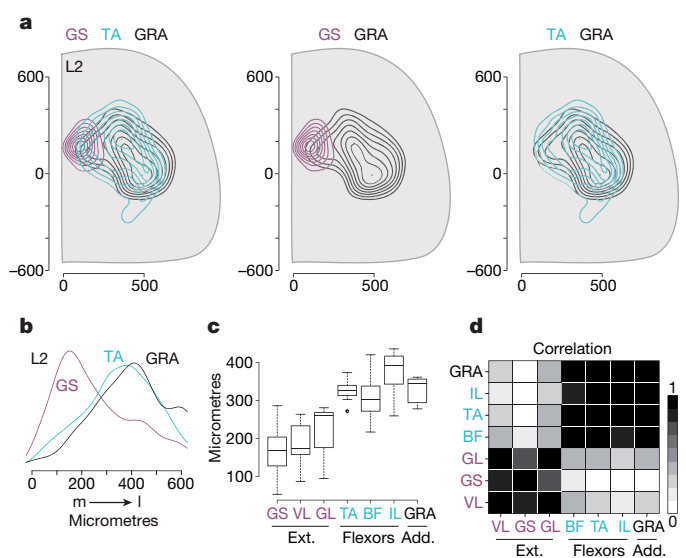


Figure 2 | Adductor and flexor premotor interneurons co-segregate. **a**, Transverse section density analysis of GS, TA and GRA premotor interneuron distributions (L2; scale in micrometres). **b**, Medio-lateral density of GS, TA and GRA premotor interneurons (L2; GRA versus TA, $P > 0.05$; GRA versus GS, $P < 0.0001$, Wilcoxon test). **c**, **d**, Medians of medio-lateral position (**c**; $P < 0.0001$, ANOVA; post-hoc analysis between all flexor–extensor pairs $P < 0.0001$, extensor–extensor or flexor–flexor $P > 0.05$; Tukey's HSD test) and cluster analysis (**d**; correlation index in arbitrary units; scale to the right) of premotor populations investigated.

GL form a separate cluster (Fig. 2d). Our data therefore support the existence of a distinct anatomical representation of interneuron groups premotor to functionally different motor neuron pools.

Dorsal dominance in spatial segregation

To narrow down the cellular source(s) responsible for the observed medio-lateral distribution differences of premotor interneurons, we divided the spinal cord into distinct domains. Medio-lateral distribution differences between extensor and flexor premotor interneurons were only observed in the ipsilateral spinal cord (Fig. 1f and Supplementary Fig. 3). Analysis of medio-lateral interneuron segregation at individual rostro-caudal levels showed that differences between GS and TA were present throughout but most pronounced at rostral lumbar levels, whereas for the VL/BF pair, medio-lateral distribution differences were smaller and only statistically significant at rostral levels (Supplementary Fig. 4). We next gated our analysis to premotor interneurons located dorsally or ventrally to the central canal, and found medio-lateral differences exclusively for interneurons with dorsal cell body position (Supplementary Fig. 5). Taken together, these findings reveal a clear medio-lateral segregation of extensor and flexor premotor interneurons in the ipsilateral dorsal spinal cord.

Molecular profile of dorsal premotor interneurons

During development, spinal interneurons emerge from distinct transcriptionally defined progenitor domains, generating six dorsally derived (d11–6) and four ventrally derived (V0–3) broad interneuron classes^{6,27}. d14–6 subclasses arise from intermediate spinal progenitor domains, express the homeodomain transcription factor *Lbx1* and encompass both excitatory and inhibitory neurons^{28,29}. Using developmental lineage tracing^{30,31} (Supplementary Fig. 6a), we found that the *Lbx1* progenitor domain contributes to dorsal extensor and flexor premotor interneurons with a distribution bias towards extensor populations (Supplementary Fig. 6b; GS, 61%; TA, 31% of all dorsal), but medio-lateral density distributions were comparable to all dorsal premotor interneurons (Supplementary Fig. 6b).

Supporting the existence of inhibitory and excitatory premotor interneurons encompassed within the *Lbx1*-marked cohort, we found co-labelling of extensor and flexor premotor interneurons in *GAD67^{GFP}* and *GlyT2^{GFP}* transgenic mice^{32,33}, as well as in *GAD65^{GFP}* Bac transgenic mice with previously described connections between GFP^{on} neurons and motor neurons³⁴ (Supplementary Fig. 7). In addition, extensor and flexor premotor interneurons also encompass *Tlx3^{on}* neurons (Supplementary Fig. 7), consistent with excitatory neurotransmitter fate³⁵. Together, these findings show that the *Lbx1* progenitor domain gives rise to dorsal premotor interneurons with an extensor versus flexor bias, but alone does not offer a clean distinguishing arbiter between these subgroups, raising the question of other sources for distinction.

Timing of neurogenesis of premotor interneurons

Birth-dating studies have provided evidence for a temporal neurogenesis gradient in the spinal cord, with the ventral-most neurons generated before dorsally derived populations³⁶. *Lbx1*-derived neurons are generated during two waves^{27,28}, raising the question of whether dorsal extensor and flexor premotor interneurons derived from this domain are generated at different developmental stages. Pulse-chase experiments with two birth-dating tags, sequentially injecting BrdU (e10.5) and EdU (e12.5), revealed clear differences along both dorso-ventral³⁶ and medio-lateral axes in p12 spinal cords, with neurons generated at e10.5 taking on a more ventral and lateral position than neurons generated at e12.5 (Fig. 3a and Supplementary Fig. 8).

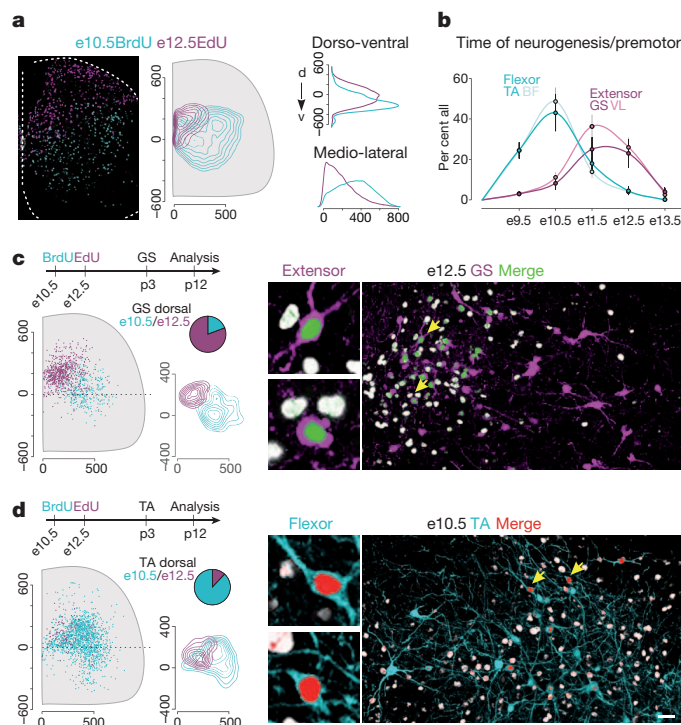


Figure 3 | Timing of neurogenesis separates extensor and flexor populations. **a**, p12 lumbar spinal cord pulsed with BrdU (e10.5) and EdU (E12.5) (left), transverse section density analysis (middle; dorsal horn excluded), and dorso-ventral and medio-lateral density analysis (right; dorsal horn excluded). **b**, Time of neurogenesis for TA, BF, GS and VL premotor interneurons (e9.5–e13.5 daily intervals; \pm s.d.). **c**, **d**, Neurogenesis analysis for GS (**c**) and TA (**d**) premotor interneurons with e10.5 (turquoise) and e12.5 (purple) birth-dating tags. Left: BrdU or EdU^{on}/rabies^{on} neurons represented as single dots (top right inset shows fractional distributions of dorsal premotor interneurons at e10.5 versus e12.5; bottom right inset shows density analysis of neuronal distributions). Right: representative examples for e12.5 (**c**) or e10.5 (**d**) pulse together with viral labelling of GS (**c**, purple) or TA (**d**, turquoise) premotor interneurons (left panels show high magnification of birth-dating tags). Scale bar: **a**, 70 μ m; **c**, **d**, 10 μ m left, 30 μ m right.

We next gated the birth-dating tags to either extensor or flexor premotor interneurons. We found enrichment for flexor premotor interneurons (Fig. 3d and Supplementary Fig. 9; TA, $43 \pm 9\%$; BF, $48 \pm 7\%$) compared to extensor premotor interneurons generated at e10.5 (Fig. 3c and Supplementary Fig. 9; GS, $8 \pm 3\%$; VL, $11 \pm 3\%$). In contrast, a selective gate to the e12.5 birth-dating tag revealed a bias for extensor premotor interneurons (Fig. 3c and Supplementary Fig. 9; GS, $23 \pm 7\%$; VL, $26 \pm 5\%$) compared to flexor populations (Fig. 3d and Supplementary Fig. 9; TA, $4 \pm 2\%$; BF, $4 \pm 1\%$). Injections at daily intervals (e9.5–e13.5) confirmed the separate neurogenesis peaks for flexor and extensor premotor interneurons (Fig. 3b).

We next determined whether similar birth-dating order rules apply to neurons derived from defined progenitor domains. We found that e10.5-generated *Lbx1*-derived flexor premotor interneurons outnumber their extensor counterparts (e10.5 TA, 23%; e10.5 GS, 4% of all LacZ^{on}/rabies^{on} neurons), and that e12.5-generated *Lbx1*-derived extensor premotor interneurons dominate over flexor subsets (e12.5 TA, 8%; e12.5 GS, 29% of all LacZ^{on}/rabies^{on} neurons). These findings provide evidence for a distinction between extensor and flexor premotor interneuron populations generated from the same progenitor domain territory based on timing of neurogenesis.

Proprioceptors target extensor premotor interneurons

Spatial segregation of extensor and flexor premotor interneurons offers an opportunity for regulation by distinct synaptic sources. Proprioceptor-driven disynaptic connections from intermediate spinal neurons to extensor motor neurons are well documented by electrophysiology^{37–40}. Because this area exhibits a seemingly similar position to the highest extensor premotor interneuron density, we assessed a possible overlap by gating our analysis to this domain. Proprioceptive afferents exhibit a major termination zone in Rexed's lamina VI, and synaptophysin–GFP (SynGFP) accumulation in proprioceptive terminals shows co-labelling of SynGFP and vGlut1 (>90%; Fig. 4a and Methods). We found that $56 \pm 6\%$ of all virus-labelled dorsal GS premotor interneuron cell bodies, but only $13 \pm 8\%$ of the dorsal TA population, resided in this area (Fig. 4b, c; L1–L3 level), and a similar divergence was observed for VL/BF premotor interneurons (VL, $58 \pm 9\%$; BF, $22 \pm 2\%$; Fig. 4c). In agreement with the observed co-segregation of GRA and flexor premotor interneuron distributions, only $23 \pm 13\%$ of the dorsal GRA premotor interneuron cell bodies resided within the proprioceptive gate (Fig. 4c). In addition to their medial enrichment, dorsal extensor premotor interneurons also showed diminished accumulation in the domain lateral to the proprioceptive gate when compared to flexor populations (2.6-fold flexor enrichment; $P < 0.05$, two-tailed unpaired *t*-test). High-resolution image analysis demonstrated that both extensor and flexor premotor interneurons resident within the proprioceptor gate are targeted directly by vGlut1^{on} terminals (Fig. 4d; appositions per $100 \mu\text{m}^2$: GS, 3.0 ± 1.2 ; TA, 2.6 ± 0.7), but at values an order of magnitude higher than in the lateral domain (appositions per $100 \mu\text{m}^2$: GS, 0.2 ± 0.2 ; TA, 0.24 ± 0.1). Together, these findings show that medio-dorsally located extensor premotor interneurons co-segregate with the intermediate spinal cord proprioceptor termination area, whereas most of the flexor and GRA premotor interneuron cell bodies are positioned more laterally.

Proprioceptor ablation perturbs connectivity

To address the question of whether proprioceptive afferents themselves are involved in controlling differentiation and/or connectivity of dorsal extensor and flexor premotor interneurons, we made use of a genetic strategy ablating proprioceptors by diphtheria toxin (*PV^{DTA}* mice)⁴¹. We found that molecular identity and neuronal settling pattern related to developmental ontogeny was not perturbed in *PV^{DTA}* mice (Supplementary Fig. 10). Medio-lateral distribution profiles of TA premotor interneurons in *PV^{DTA}* mice revealed no significant differences to wild type, but marked changes were observed for GS premotor

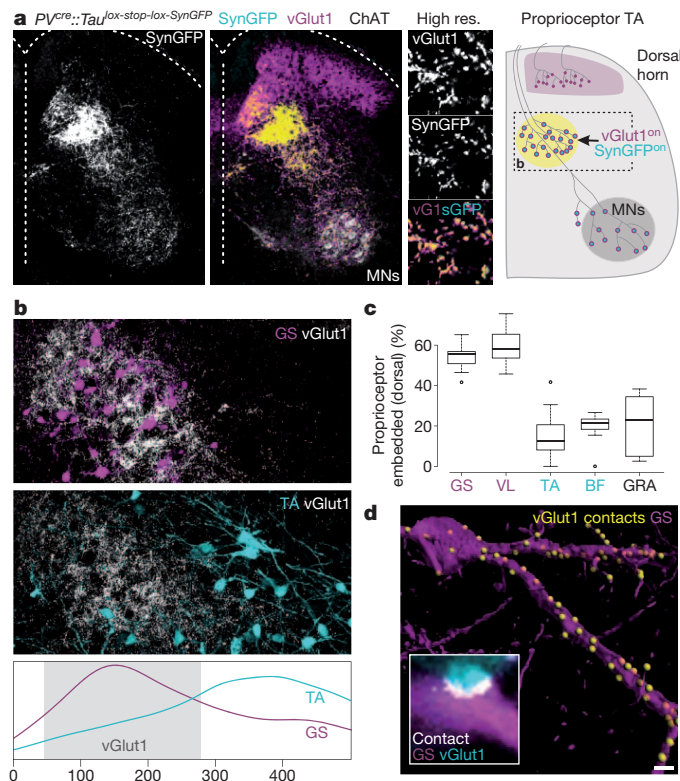


Figure 4 | Dorsal extensor premotor interneurons reside in proprioceptive termination area. **a**, Proprioceptive afferent (PA) termination zones in p12 *PV^{cre}/Tau^{lox-stop-lox-SynGFP}* lumbar spinal cord (motor neurons (MNs): ChAT^{on}). SynGFP marks PV^{on} proprioceptor terminals (left) co-localizing with vGlut1 (right; high resolution). Diagram (right): proprioceptor termination zones in intermediate (yellow) and ventral (grey; MNs) spinal cord (dorsal horn terminating afferents: vGlut1^{on}/SynGFP^{on}). **b**, GS and TA premotor interneuron analysis gated to intermediate spinal cord vGlut1^{on} PA termination zone (white). The bottom panel shows medio-lateral density distribution of GS/TA premotor interneurons (grey, medio-lateral extent of vGlut1^{on} termination zone). **c**, Percentage of proprioceptor-embedded dorsal GS, VL, TA, BF and GRA premotor interneurons ($P < 0.0001$, ANOVA; post-hoc analysis between all pairs of flexor–extensor $P < 0.0001$, extensor–extensor or flexor–flexor $P > 0.05$; Tukey's HSD test). **d**, Synaptic contacts between vGlut1^{on} terminals (yellow: reconstructed appositions) and GS premotor interneuron. Inset: IMARIS analysis (turquoise, vGlut1; white, apposition vGlut1/GS premotor interneuron). Scale bars: **a**, 70 μm (left), 4 μm (right); **b**, 20 μm ; **d**, 4 μm (inset, 0.5 μm).

interneurons (Fig. 5a, b, d, f). Whereas the highest overall density of GS premotor interneurons in wild type was found medially, a pronounced lateral shift was observed in *PV^{DTA}* mice, resulting in a density distribution profile indistinguishable from flexor (Fig. 5d, f). Unbiased monitoring of GS and TA populations by cluster analysis revealed that the *PV^{DTA}* GS population co-segregates with the wild-type TA rather than GS population (Fig. 5g), a phenotype evident throughout the rostro-caudal extent for both GS/TA and VL/BF pairs (Supplementary Fig. 11). By application of digital masks to the area normally occupied by proprioceptive afferents in wild-type mice, we found that only $6 \pm 4\%$ of all dorsal GS interneurons in *PV^{DTA}* mice were located in this now proprioceptor-stripped domain, compared to $56 \pm 6\%$ in wild type (Fig. 5e).

Group Ia proprioceptive afferents also establish synaptic contacts with motor neurons, in addition to their contacts in the intermediate spinal cord. To determine whether contact by proprioceptors to the intermediate spinal cord is sufficient to implement motoneuronal connections by extensor premotor interneurons, we analysed *Er81* mutant mice, which have proprioceptive connectivity to intermediate spinal cord but not motor neurons⁴². *Er81* mutant mice revealed no

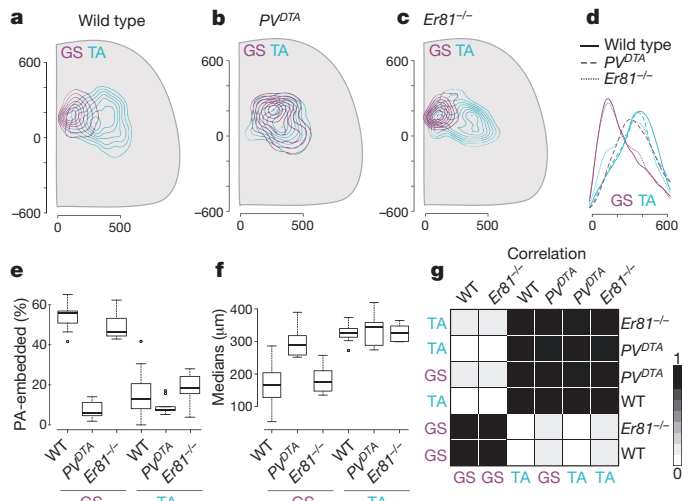


Figure 5 | Proprioceptors control connectivity of extensor premotor interneurons. **a–d**, Transverse (**a–c**) and medio-lateral (**d**) density analysis of GS and TA premotor interneurons in wild-type (**a**, **d**), *PV^{DTA}* (**b**, **d**) and *Er81^{-/-}* (**c**, **d**) mice. Scale in micrometres. **e**, Percentage of dorsal GS/TA premotor interneuron cell bodies within proprioceptive afferent (PA) termination area ($P < 0.0001$, ANOVA; post-hoc analysis; GS: *PV^{DTA}* versus wild type $P < 0.0001$, *Er81^{-/-}* versus wild type $P > 0.6$; TA: *PV^{DTA}* versus wild type $P > 0.3$, *Er81^{-/-}* versus wild type $P > 0.7$, Tukey's HSD test). **f**, **g**, Medio-lateral position (**f**; $P < 0.0001$, ANOVA; post-hoc analysis, GS: *PV^{DTA}* versus wild type $P < 0.0001$, *Er81^{-/-}* versus wild-type $P > 0.8$; TA: *PV^{DTA}* versus wild-type $P > 0.9$, *Er81^{-/-}* versus wild-type $P > 0.9$, Tukey's HSD test) and cluster analysis of median distributions (**g**; correlation index in arbitrary units; scale to the right) of GS/TA premotor interneurons in wild-type, *PV^{DTA}* and *Er81^{-/-}* mice.

significant difference in the medio-lateral density distribution profiles of GS extensor premotor interneurons compared to wild-type mice (Fig. 5a, c, d, f, g and Supplementary Fig. 11). Together, these findings demonstrate that the establishment of direct connections between presumptive extensor 'premotor' interneurons in the medio-dorsal spinal cord is dependent on the presence of proprioceptors in the intermediate spinal cord.

Discussion

Motor antagonism must be reflected at the neuronal circuit level, supporting initiation, propagation and alternation of movement based on the changing needs of the system. Our study provides insight into the organizational principles of motor output circuitry at the global premotor level, just one synapse away from motor neurons. This approach has the previously unmet potential to lay open differential anatomical three-dimensional connectivity matrices to functionally defined motor neuron pools, and exposes premotor network organization indiscriminate of functional criteria such as neuronal activity patterns. More refined analysis of premotor cohorts will reveal which neuronal subpopulations can be assigned to specific expressions of motor antagonism or other motor circuit modules. For instance, although rhythmically active interneurons can be observed in the intermediate and ventral spinal cord in adult cats⁴³ and motor burst generation in neonatal rodents is thought to occur ventrally⁴⁴, the relationship between the premotor network described here and antagonistic central pattern generator modules^{2,10,11} awaits future investigations.

We found the most marked organizational differences between overall extensor and flexor premotor distributions in the intermediate spinal cord with pronounced sensory input, indicating that spatial organization here may be important for the function of premotor circuits involved in sensory-regulated antagonism. Proprioceptive feedback provides an essential arbiter intersecting with spinal circuits

to adjust motor output^{15,45}. Interestingly, a disynaptic circuit acting on lamina V/VI interneurons switching between inhibition at rest and excitation during extension has been proposed based on functional studies in adult cat^{37–40}. Our findings on dorso-medial extensor premotor interneurons provide a possible anatomical substrate for these last-order interneurons as targets for more distant circuit components. They reveal the existence of an evolutionarily conserved disynaptic circuit in mammals, in which proprioceptors act as a ‘private synaptic channel’ to spatially segregated premotor interneuron populations. During development, proprioceptive afferents control monosynaptic connectivity between cognate dorsal extensor premotor interneurons and motor neurons. Whether other synaptic sources such as defined descending pathways have analogous roles in providing input to and influencing connection specificity of lateral extensor and flexor populations—imposed by spatial organizational constraints—is an interesting question to pursue.

Focusing on premotor spinal interneurons, we uncover a spatial and developmental motor map at the level of extensor–flexor premotor circuits. Our findings provide support for an anatomy-based organization of premotor circuits, together with the observation that circuits in larval zebrafish are anatomically tuned to motor function during development with correlation between dorso-ventral topographic cell body organization and movements at changing speed⁴⁶. We found that the medio-lateral extensor–flexor premotor interneuron segregation in the dorsal spinal cord has its origin in temporally divergent neuronal generation. Specification of temporal neuronal identity is best studied in *Drosophila*, where neuroblast generation in a temporal sequence is accompanied by transcription factor cascades⁴⁷. In mammals, a correlation between temporal identity and connectivity has been observed in cerebellar, hippocampal and olfactory circuits^{48–50}, but whether and how these differences relate to function and behaviour remains unclear. Previous developmental and genetic studies aimed at understanding circuits underlying functional antagonism in mice were largely based on the dorso-ventral progenitor domain origin of neuronal populations^{6,7}. Our study suggests that temporal identity contributes an essential parameter to the identification of the circuit basis underlying functional dichotomy in the mammalian spinal cord and provides insight into the organizational principles of neuronal circuits controlling motor behaviour.

METHODS SUMMARY

For the generation of *Tau^{lox-stop-lox-SynGFP}* mice, a *lox-stop-lox-synaptophysin-GFP-Nls-lacZ-pA* targeting cassette was integrated into exon 2 of the *Tau* locus as previously described³¹. Other mouse strains and intercrosses are described in Methods. Virus production and injections were performed as described^{22–24}. Images were acquired with an Olympus confocal (FV1000) or a custom-made dual spinning-disk microscope²⁴. Proprioceptive gates were assigned based on vGlut1 analysis in Matlab, where a pixel intensity-based mask was assigned to each acquired image to delimitate the gate borders.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 April; accepted 2 September 2011.

Published online 19 October 2011.

1. Brown, T. G. The intrinsic factor in the act of progression in the mammal. *Proc. R. Soc. Lond.* **84**, 308–319 (1911).
2. Grillner, S. Biological pattern generation: the cellular and computational logic of networks in motion. *Neuron* **52**, 751–766 (2006).
3. Sherrington, C. S. Flexion-reflex of the limb, crossed extension-reflex, and reflex stepping and standing. *J. Physiol.* **40**, 28–121 (1910).
4. Stuart, D. G. & Hultborn, H. Thomas Graham Brown (1882–1965), Anders Lundberg (1920–), and the neural control of stepping. *Brain Res. Rev.* **59**, 74–95 (2008).
5. Yakovenko, S., Mushahwar, V., VanderHorst, V., Holstege, G. & Prochazka, A. Spatiotemporal activation of lumbosacral motoneurons in the locomotor step cycle. *J. Neurophysiol.* **87**, 1542–1553 (2002).
6. Goulding, M. Circuits controlling vertebrate locomotion: moving in a new direction. *Nature Rev. Neurosci.* **10**, 507–518 (2009).
7. Kiehn, O. Development and functional organization of spinal locomotor circuits. *Curr. Opin. Neurobiol.* **21**, 100–109 (2011).

8. Wang, Z., Li, L., Goulding, M. & Frank, E. Early postnatal development of reciprocal Ia inhibition in the murine spinal cord. *J. Neurophysiol.* **100**, 185–196 (2008).
9. Grillner, S. & Wallen, P. Central pattern generators for locomotion, with special reference to vertebrates. *Annu. Rev. Neurosci.* **8**, 233–261 (1985).
10. Stein, P. S. Motor pattern deletions and modular organization of turtle spinal cord. *Brain Res. Rev.* **57**, 118–124 (2008).
11. Bizzi, E., Cheung, V. C., d’Avella, A., Saltiel, P. & Tresch, M. Combining modules for movement. *Brain Res. Rev.* **57**, 125–133 (2008).
12. McCrea, D. A. & Rybak, I. A. Organization of mammalian locomotor rhythm and pattern generation. *Brain Res. Rev.* **57**, 134–146 (2008).
13. Jankowska, E. Interneuronal relay in spinal pathways from proprioceptors. *Prog. Neurobiol.* **38**, 335–378 (1992).
14. Rossignol, S., Dubuc, R. & Gossard, J. P. Dynamic sensorimotor interactions in locomotion. *Physiol. Rev.* **86**, 89–154 (2006).
15. Windhorst, U. Muscle proprioceptive feedback and spinal networks. *Brain Res. Bull.* **73**, 155–202 (2007).
16. Schouenborg, J. Modular organisation and spinal somatosensory imprinting. *Brain Res. Rev.* **40**, 80–91 (2002).
17. Eccles, J. C., Eccles, R. M. & Lundberg, A. The convergence of monosynaptic excitatory afferents on to many different species of alpha motoneurons. *J. Physiol.* **137**, 22–50 (1957).
18. Mears, S. C. & Frank, E. Formation of specific monosynaptic connections between muscle spindle afferents and motoneurons in the mouse. *J. Neurosci.* **17**, 3128–3135 (1997).
19. Lloyd, D. P. Integrative pattern of excitation and inhibition in two-neuron reflex arcs. *J. Neurophysiol.* **9**, 439–444 (1946).
20. Hultborn, H., Jankowska, E. & Lindstrom, S. Recurrent inhibition of interneurons monosynaptically activated from group Ia afferents. *J. Physiol.* **215**, 613–636 (1971).
21. Eccles, R. M. & Lundberg, A. Integrative pattern of Ia synaptic actions on motoneurons of hip and knee muscles. *J. Physiol.* **144**, 271–298 (1958).
22. Marshel, J. H., Mori, T., Nielsen, K. J. & Callaway, E. M. Targeting single neuronal networks for gene expression and cell labeling *in vivo*. *Neuron* **67**, 562–574 (2010).
23. Wickersham, I. R. et al. Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* **53**, 639–647 (2007).
24. Stepien, A. E., Tripodi, M. & Arber, S. Monosynaptic rabies virus reveals premotor network organization and synaptic specificity of cholinergic partition cells. *Neuron* **68**, 456–472 (2010).
25. McHanwell, S. & Biscoe, T. J. The localization of motoneurons supplying the hindlimb muscles of the mouse. *Philos. Trans. R. Soc. Lond.* **293**, 477–508 (1981).
26. Pratt, C. A., Chanaud, C. M. & Loeb, G. E. Functionally complex muscles of the cat hindlimb. IV. Intramuscular distribution of movement command signals and cutaneous reflexes in broad, bifunctional thigh muscles. *Exp. Brain Res.* **85**, 281–299 (1991).
27. Helms, A. W. & Johnson, J. E. Specification of dorsal spinal cord interneurons. *Curr. Opin. Neurobiol.* **13**, 42–49 (2003).
28. Gross, M. K., Dottori, M. & Goulding, M. Lbx1 specifies somatosensory association interneurons in the dorsal spinal cord. *Neuron* **34**, 535–549 (2002).
29. Müller, T. et al. The homeodomain factor Lbx1 distinguishes two major programs of neuronal differentiation in the dorsal spinal cord. *Neuron* **34**, 551–562 (2002).
30. Sieber, M. A. et al. Lbx1 acts as a selector gene in the fate determination of somatosensory and viscerosensory relay neurons in the hindbrain. *J. Neurosci.* **27**, 4902–4909 (2007).
31. Hippenmeyer, S. et al. A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS Biol.* **3**, e159 (2005).
32. Zeilhofer, H. U. et al. Glycinergic neurons expressing enhanced green fluorescent protein in bacterial artificial chromosome transgenic mice. *J. Comp. Neurol.* **482**, 123–141 (2005).
33. Kaneko, K. et al. Noradrenergic excitation of a subpopulation of GABAergic cells in the basolateral amygdala via both activation of nonselective cationic conductance and suppression of resting K⁺ conductance: a study using glutamate decarboxylase 67-green fluorescent protein knock-in mice. *Neuroscience* **157**, 781–797 (2008).
34. Wilson, J. M., Blagovetchenski, E. & Brownstone, R. M. Genetically defined inhibitory neurons in the mouse spinal cord dorsal horn: a possible source of rhythmic inhibition of motoneurons during fictive locomotion. *J. Neurosci.* **30**, 1137–1148 (2010).
35. Cheng, L. et al. *Tlx3* and *Tlx1* are post-mitotic selector genes determining glutamatergic over GABAergic cell fates. *Nature Neurosci.* **7**, 510–517 (2004).
36. Nornes, H. O. & Carry, M. Neurogenesis in spinal cord of mouse: an autoradiographic analysis. *Brain Res.* **159**, 1–6 (1978).
37. McCrea, D. A. Neuronal basis of afferent-evoked enhancement of locomotor activity. *Ann. NY Acad. Sci.* **860**, 216–225 (1998).
38. Pearson, K. G., Misiaszek, J. E. & Fouad, K. Enhancement and resetting of locomotor activity by muscle afferents. *Ann. NY Acad. Sci.* **860**, 203–215 (1998).
39. Conway, B. A., Hultborn, H. & Kiehn, O. Proprioceptive input resets central locomotor rhythm in the spinal cat. *Exp. Brain Res.* **68**, 643–656 (1987).
40. Angel, M. J., Jankowska, E. & McCrea, D. A. Candidate interneurons mediating group I disynaptic EPSPs in extensor motoneurons during fictive locomotion in the cat. *J. Physiol.* **563**, 597–610 (2005).
41. Vrieseling, E. & Arber, S. Target-induced transcriptional control of dendritic patterning and connectivity in motor neurons by the ETS gene *Pea3*. *Cell* **127**, 1439–1452 (2006).

42. Arber, S., Ladle, D. R., Lin, J. H., Frank, E. & Jessell, T. M. ETS gene *Er81* controls the formation of functional connections between group Ia sensory afferents and motor neurons. *Cell* **101**, 485–498 (2000).
43. Baev, K. V., Degtiarenko, A. M., Zavadskaya, T. V. & Kostyuk, P. G. Activity of interneurons of the lumbar region of the spinal cord during fictive locomotion of thalamic cats. *Neirofiziologiya* **11**, 329–338 (1979).
44. Kjaerulff, O. & Kiehn, O. Distribution of networks generating and coordinating locomotor activity in the neonatal rat spinal cord *in vitro*: a lesion study. *J. Neurosci.* **16**, 5777–5794 (1996).
45. Hiebert, G. W. & Pearson, K. G. Contribution of sensory feedback to the generation of extensor activity during walking in the decerebrate cat. *J. Neurophysiol.* **81**, 758–770 (1999).
46. McLean, D. L., Fan, J., Higashijima, S., Hale, M. E. & Fetcho, J. R. A topographic map of recruitment in spinal cord. *Nature* **446**, 71–75 (2007).
47. Isshiki, T., Pearson, B., Holbrook, S. & Doe, C. Q. *Drosophila* neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny. *Cell* **106**, 511–521 (2001).
48. Zong, H., Espinosa, J. S., Su, H. H., Muzumdar, M. D. & Luo, L. Mosaic analysis with double markers in mice. *Cell* **121**, 479–492 (2005).
49. Deguchi, Y., Donato, F., Galimberti, I., Cabuy, E. & Caroni, P. Temporally matched subpopulations of selectively interconnected principal neurons in the hippocampus. *Nature Neurosci.* **14**, 495–504 (2011).
50. Imamura, F., Ayoub, A. E., Rakic, P. & Greer, C. A. Timing of neurogenesis is a determinant of olfactory circuitry. *Nature Neurosci.* **14**, 331–337 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to M. Sigrist for the generation of *Tau^{lox-stop-lox-SyngFP}* mice, M. Mielich for help in the production of viruses, P. Schwarb, L. Gelman, A. Ponti and M. Stadler for help and advice with image acquisition and statistical analysis, and to P. Caroni, G. Courtine, T. Jessell, B. Roska and P. Scheiffele for discussions and comments on the manuscript. We thank E. Callaway and K. Conzelmann for advice on virus work, and C. Birchmeier for *Lbx1^{Cre}* mice and Tlx3 antibodies. M.T. was supported by an EMBO long-term fellowship, M.T., A.E.S. and S.A. by an ERC Advanced Grant, the Swiss National Science Foundation, the Kanton Basel-Stadt, EU Framework Program 7 and the Novartis Research Foundation.

Author Contributions M.T. carried out experiments and data acquisition, and was involved in design of experiments, data analysis and writing of the manuscript. A.E.S. developed the trans-synaptic virus method applied in this study. S.A. initiated the project and design of experiments, analysed data and wrote the manuscript. All authors discussed the experiments and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.A. (silvia.arber@unibas.ch).

METHODS

Generation of mice and mouse genetics. *PV^{cre}* (ref. 31), *Tau^{lox-stop-lox-mGFP}* (ref. 31), *Lbx1^{cre}* (ref. 30), *GAD65^{GFP}* (ref. 34), *GAD67^{GFP}* (ref. 33), *GlyT2^{GFP}* (ref. 32), *Isl2^{DTA}* (ref. 51) and *Er81* mutant⁴² mouse strains have been described previously. Intersectional breeding of *PV^{cre}* (ref. 31) and *Isl2^{DTA}* (ref. 51) mice was used to selectively ablate proprioceptors in this study. At early postnatal stages, these mice lack vGlut1^{on} signal in the intermediate and ventral spinal cord^{41,52} as well as peripheral end organs normally innervated by proprioceptors⁵², demonstrating the complete absence of proprioceptive afferents. Wild-type mice and intercrosses were maintained on a mixed genetic background (129/C57BL/6), and all animal experiments were performed according to Swiss guidelines.

Virus tracing experiments and timing of neurogenesis. The approach used in this study makes use of an attenuated rabies virus expressing fluorescent marker proteins (mCherry or eGFP) exhibiting a selective deletion in a glycoprotein (Gly) essential for trans-synaptic spreading^{22,23}. Targeted co-injection of Δ Gly-rabies virus (Rab-mCherry or Rab-eGFP) and an adeno-associated virus expressing Gly (AAV-Gly) into limb muscles results in retrograde infection and labelling of premotor neurons in addition to spinal motor neuron pools innervating the injected muscles²⁴. Motor neurons are located in Rexed's lamina IX and can be distinguished from interneurons on the basis of choline acetyl-transferase (ChAT) expression²⁴. Rab-mCherry, Rab-GFP and AAV-Gly viruses were produced and titrated as described²⁴. Muscles were identified and assigned according to ref. 53, and injection specificity for every experiment was confirmed using fluorescence upon humanely killing the animals.

BrdU (Sigma)/EdU (Invitrogen) were injected at 100 mg kg⁻¹ in pregnant female mice at the desired embryonic stages. For BrdU experiments, mice were perfusion fixed and subsequently immersion fixed for 4 h in 4% PFA/0.8% picric acid. Tissue-Tek OCT embedded spinal cords were cut at 40 μ m using a cryostat. Floating sections were incubated for 30 min at 45 °C in 1 M HCl, briefly washed in PBS, incubated for 10 min in borate buffer pH 8.0 followed by another PBS washing step. After this treatment, EdU was revealed as described⁵⁴ (Invitrogen Click-it EdU). After EdU staining, mouse anti-BrdU was added over night at 4 °C and revealed with fluorescent-conjugated secondary antibodies (Invitrogen).

Rationale for choice of muscles for premotor interneuron analysis. The following muscles were chosen for analysis: tibialis anterior (TA), posterior biceps femoris (BF), iliopsoas (IL), gastrocnemius (GS), vastus lateralis (VL), gluteus (GL) and gracilis (GRA). These muscles were selected based on the following criteria: (1) the mechanical actions of the muscles on joints, taking into account the anatomical muscle attachment sites through tendons; (2) the reflex response patterns as defined previously³, providing a comprehensive reflex analysis in the cat, and defining flexor muscles as activated and extensor muscles as unaffected or inhibited by flexor reflex afferent stimulation; (3) the phase of muscle recruitment during forward locomotion, based on EMG data in *in vivo* preparations, preferentially in mice or rats, but referring to cat data when such data are not available from mice or rats. As defined in our study, extensor muscles (GS, VL, GL) act mechanically as extensors, are inhibited or unaffected during the flexion reflex and are recruited during stance phase in forward locomotion *in vivo*, whereas flexor muscles (TA, BF, IL) act mechanically as flexors, are activated during the flexion reflex and are recruited during swing phase in forward locomotion *in vivo*.

In more detail, GS acts as an ankle extensor, is recruited during stance in mice *in vivo*⁵⁵ and is inhibited during the flexion reflex in cat³. VL acts as a knee extensor, is recruited during stance in mice *in vivo*⁵⁵, and is inhibited during the flexion reflex in cat³. GL acts as a hip extensor, is recruited during stance phase in forward locomotion *in vivo*⁵⁶ and is unaffected by the flexion reflex in cat³. TA acts as ankle flexor, is recruited during swing phase in forward locomotion in mice *in vivo*⁵⁵ and is excited during the flexion reflex in cat³. The choice of BF is more complex owing to the bi-functional nature of this muscle, but its choice was obligatory due to the absence of pure knee flexor muscles in mice. For this reason, we chose to inject the posterior head of the BF, which inserts to the distal shank, and as a consequence, this head has a prominent mechanical role in knee flexion, with a hip-to-knee lever arm ratio of 3:5 (ref. 57). Consistent with the more prominent role of this head in knee flexion, it is active during swing²⁶ and it is excited during the flexion reflex³. IL acts as hip flexor, is recruited during swing phase in forward locomotion *in vivo*^{56,58} and is excited during the flexion reflex³. GRA was chosen as an example of a muscle with a more complex function. Albeit acting mechanically as knee flexor, it also acts as an adductor of the thigh³. GRA therefore shows a mechanical function distinct from all other muscles examined in this study. GRA is consistently recruited during swing phase in forward locomotion *in vivo*²⁶ and is excited during the flexion reflex³, but a more complex reflex response pattern has also been reported²⁶. With respect to locomotor cycle recruitment phase, BF and GRA have also been suggested to form a separate group as retractors, muscles active to complete the backward movement of foot and toes and to lift the toes⁵.

Our analysis and choice of muscles also included a component of assessing the developmental origin of a limb muscle and columnar identity of motor neurons innervating the muscle. Genetic programmes guide the establishment of axonal trajectories of motor neurons residing in the medial and lateral aspect of the lateral motor column (LMCm, ventral limb projections; LMCl, dorsal limb projections) to muscles of distinct developmental origin^{59,60}. For instance, GS and BF are muscles of ventral developmental origin and are innervated by LMCm motor neurons, whereas TA and VL are muscles of dorsal origin and innervated by LMCl motor neurons^{25,60}. The study of premotor interneurons connected to these motor neuron pools therefore also allows addressing whether the observed density segregation is linked to the extensor and flexor function or the developmental origin of the muscle.

Immunohistochemistry and imaging. Antibodies used in this study were: chicken anti-GFP (Invitrogen), guinea-pig anti-Tlx3 (ref. 61) (gift from C. Birchmeier, Berlin), guinea-pig anti-vGlut1 (Chemicon), goat anti-ChAT (Chemicon), goat anti-LacZ (Biogenesis), mouse anti-BrdU (Invitrogen, MoBU-1 clone), rabbit anti-GFP (Invitrogen), rabbit anti-LacZ (Invitrogen), rabbit anti-PV (Swant) and rabbit anti-RFP (Chemicon). Spinal cords were sectioned and processed for immunohistochemistry as previously described²⁴ with the exception of BrdU experiments (see above).

In the ventral spinal cord, vesicular glutamate transporter 1 (vGlut1) identifies synaptic terminals of proprioceptors⁶². Mechanoreceptive sensory neurons with deep dorsal horn projections also express vGlut1 (ref. 63), prompting us to determine whether vGlut1 faithfully marks proprioceptive terminals in the intermediate spinal cord proprioceptor termination area. Expression of the fusion protein synaptophysin-GFP (Syn-GFP) in proprioceptors of *PV^{cre}/Tau^{lox-stop-lox-SynGFP}* mice highlights the proprioceptor termination area in the intermediate spinal cord, in which >90% of terminals were co-labelled for SynGFP and vGlut1 at p12 (Fig. 5a).

Synaptic input (vGlut1^{on}) to dorsal GS and TA premotor interneurons was reconstructed using IMARIS software, for neuronal populations residing inside or outside the proprioceptive gate. Images were deconvolved using Huygens. Neuronal surfaces were reconstructed using the IMARIS surface tool, including the unambiguously assigned dendrites, before the overall surface area was calculated for each reconstructed neuron. Appositions between vGlut1^{on} terminals and premotor interneurons were determined by co-localization analysis in IMARIS, and resulting assigned synaptic contacts were extracted and counted in IMARIS using the Spots function. Values in the text express the number of appositions per 100 μ m² of reconstructed neuronal surface. For each sampled class (TA_{inside}, TA_{outside}, GS_{inside}, GS_{outside}), synaptic input to ≥ 30 neurons was determined.

Statistics. The custom-made MATLAB plug-in 'Reference Axes' was developed to reconstruct the position of virally labelled neurons in a three-dimensional Cartesian space, allowing the comparison of multiple specimens. The plug-in runs in the previously developed MATLAB-based image processing suite 'Qu'. In the Reference axes plug-in, a set of Cartesian axes is constructed interactively over all frames of a z-series with the zero centred on the central canal, the y axis parallel to the midline of the spinal cord and the x axis orthogonal to it. The position of virally labelled interneurons detected by means of Qu's spot detection algorithm is transformed (translation and rotation) to a common coordinate system. Corrected spot positions can be further processed in Qu or exported as comma-separated value files. All statistical analysis and data set visualization were performed using R (R Foundation for Statistical Computing, <http://www.r-project.org>). One-dimensional kernel density estimates were obtained using the R 'density' function. Two-dimensional kernel density estimation used to compute the distribution contours was obtained using the 'kde2d' function provided in the 'MASS' library. Two-dimensional kernel density estimation were graphically displayed as contour plots, with the contour lines connecting points of equal densities and drawn for density values between 30% and 100% of the estimated density range, in steps of 10%. Correlation matrices were computed as a function of the distances between distribution medians. For all box plots shown, the horizontal bar inside the box indicates the median value, box limits are set at the 25th and 75th percentile of the distribution, whiskers extend to accommodate values up to 1.5 times the interquartile range, values outside this range are plotted as single points. For linear density distributions, integral of curve always equals 1, and figure panels do not display y-axis values. To determine the relative ratio of extensor to flexor premotor interneurons in the dorsal spinal cord lateral to the proprioceptive gate, connectivity indices were normalized to the number of contralateral premotor interneurons, a population without any obvious extensor–flexor bias. Flexor premotor interneurons were significantly enriched (2.6 \times) when compared to their extensor counterparts (GS, $2.4 \pm 1.3 \times$ GS_{contra}; TA, $6.2 \pm 1.2 \times$ TA_{contra}; $P < 0.05$, two-tailed unpaired t-test) in wild-type mice.

51. Yang, X. *et al.* Patterning of muscle acetylcholine receptor gene expression in the absence of motor innervation. *Neuron* **30**, 399–410 (2001).

52. Friesse, A. *et al.* Gamma and alpha motor neurons distinguished by expression of transcription factor Err3. *Proc. Natl Acad. Sci. USA* **106**, 13588–13593 (2009).
53. Greene, E. C. *Anatomy of the Rat* (Hafner, 1935).
54. Chehrehasa, F., Meedeniya, A. C., Dwyer, P., Abrahamsen, G. & Mackay-Sim, A. EdU, a new thymidine analogue for labelling proliferating cells in the nervous system. *J. Neurosci. Methods* **177**, 122–130 (2009).
55. Akay, T., Acharya, H. J., Fouad, K. & Pearson, K. G. Behavioral and electromyographic characterization of mice lacking EphA4 receptors. *J. Neurophysiol.* **96**, 642–651 (2006).
56. Carrier, L., Brustein, E. & Rossignol, S. Locomotion of the hindlimbs after neurectomy of ankle flexors in intact and spinal cats: model for the study of locomotor plasticity. *J. Neurophysiol.* **77**, 1979–1993 (1997).
57. Chanaud, C. M. & Macpherson, J. M. Functionally complex muscles of the cat hindlimb. III. Differential activation within biceps femoris during postural perturbations. *Exp. Brain Res.* **85**, 271–280 (1991).
58. de Leon, R. D., Tamaki, H., Hodgson, J. A., Roy, R. R. & Edgerton, V. R. Hindlimb locomotor and postural training modulates glycinergic inhibition in the spinal cord of the adult spinal cat. *J. Neurophysiol.* **82**, 359–369 (1999).
59. Tsuchida, T. *et al.* Topographic organization of embryonic motor neurons defined by expression of LIM homeobox genes. *Cell* **79**, 957–970 (1994).
60. Jones, C. L. The morphogenesis of the thigh of the mouse with special reference to tetrapod muscle homologies. *J. Morphol.* **162**, 275–309 (1979).
61. Muller, T. *et al.* The bHLH factor Olig3 coordinates the specification of dorsal neurons in the spinal cord. *Genes Dev.* **19**, 733–743 (2005).
62. Pecho-Vrieseling, E., Sigrist, M., Yoshida, Y., Jessell, T. M. & Arber, S. Specificity of sensory–motor connections encoded by Sema3e–Plxnd1 recognition. *Nature* **459**, 842–846 (2009).
63. Luo, W., Enomoto, H., Rice, F. L., Milbrandt, J. & Ginty, D. D. Molecular identification of rapidly adapting mechanoreceptors and their developmental dependence on ret signaling. *Neuron* **64**, 841–856 (2009).

Melanopsin signalling in mammalian iris and retina

T. Xue^{1,2}, M. T. H. Do^{1,2,†}, A. Riccio³, Z. Jiang^{1,2}, J. Hsieh^{1,4}, H. C. Wang^{1,5}, S. L. Merbs⁶, D. S. Welsbie⁶, T. Yoshioka^{1,7}, P. Weissgerber⁸, S. Stolz⁸, V. Flockerzi⁸, M. Freichel^{8,†}, M. I. Simon⁹, D. E. Clapham³ & K.-W. Yau^{1,2,6}

Non-mammalian vertebrates have an intrinsically photosensitive iris and thus a local pupillary light reflex (PLR). In contrast, it is thought that the PLR in mammals generally requires neuronal circuitry connecting the eye and the brain. Here we report that an intrinsic component of the PLR is in fact widespread in nocturnal and crepuscular mammals. In mouse, this intrinsic PLR requires the visual pigment melanopsin; it also requires PLC β 4, a vertebrate homologue of the *Drosophila* NorpA phospholipase C which mediates rhabdomic phototransduction. The *Plcb4*^{-/-} genotype, in addition to removing the intrinsic PLR, also essentially eliminates the intrinsic light response of the M1 subtype of melanopsin-expressing, intrinsically photosensitive retinal ganglion cells (M1-ipRGCs), which are by far the most photosensitive ipRGC subtype and also have the largest response to light. Ablating in mouse the expression of both TRPC6 and TRPC7, members of the TRP channel superfamily, also essentially eliminated the M1-ipRGC light response but the intrinsic PLR was not affected. Thus, melanopsin signalling exists in both iris and retina, involving a PLC β 4-mediated pathway that nonetheless diverges in the two locations.

The discovery of ipRGCs has overturned the century-old belief that rods and cones are the only mammalian retinal photoreceptors^{1–6}. These ganglion-cell photoreceptors serve primarily non-image visual functions, one of which is the PLR. In animals such as fish, amphibians and bird, in addition to the neurally driven PLR, the iris itself is capable of autonomous, light-induced constriction^{7–11}. For mammals, it is thought that the PLR generally requires neuronal circuitry through the brain, although sporadic reports^{7,12,13} and some controversy exist of an intrinsic iridic photosensitivity in occasional species, including human. Even in non-mammalian vertebrates, the photopigment driving the intrinsic PLR remains unidentified. In amphibians and fish this photopigment has been proposed to be rhodopsin^{8,9}, whereas in chicken it is suggested to be the non-opsin-based cryptochrome¹¹.

We have examined the above questions more closely, and found that the intrinsic PLR is surprisingly widespread in mammals. Moreover, it bears a close relationship to the ipRGCs in phototransduction.

Intrinsic PLR in mouse and other mammals

We found that bright light triggered a pupillary constriction in an intact eye freshly isolated from a dark-adapted pigmented mouse (Fig. 1a; $n = 3$ eyes; Supplementary Methods). This photosensitivity disappeared within ~ 1 min after eye isolation, perhaps due to anoxia, with the pupil remaining constricted thereafter. This PLR persisted in a reduced preparation with just the eye's anterior chamber and iris^{8,11,12} under L-15 medium (Fig. 1b; $n = 3$ eyes; Supplementary Methods), even after blocking any potential parasympathetic activity to the iris with 0.5% atropine; however, the PLR again faded after a few light trials.

The mammalian iris has three main tissue layers, all pigmented with melanin: an anterior fibrovascular stroma; a middle smooth-muscle

layer consisting of the circumferential sphincter muscle at the pupil perimeter and the radial dilator muscle across the iris; and a posterior epithelium^{7,14}. In non-mammalian vertebrates, the sphincter muscle itself is thought to be the light sensor^{7,14}. Accordingly, we reduced the mouse preparation further to the small ring of sphincter muscle, and connected it to a strain gauge for measuring isometric tension under oxygenated Ames solution^{9,10} (35–37 °C; Supplementary Methods). The isolated sphincter muscle reproducibly gave a light-induced contractile force for hours. A relatively dim flash elicited a transient increase in force that grew linearly with increasing flash intensity; that is, proportional to flash intensity and with a constant waveform (Fig. 1c and inset, bottom; see also Supplementary Fig. 1). This flash-induced contraction reached a transient peak in ~ 1 s but decayed much more slowly, not dissimilar to the *in situ* ipRGC-driven PLR¹⁵. The response to an intense flash often showed a hump during its decay (Fig. 1c inset, top). Although the force elicited by a dim flash decayed to baseline in ≤ 1 min, a second identical flash typically elicited a smaller response unless given in ≥ 8 min after the first. Likewise, a near-saturated response to an intense flash decayed in ~ 1 min, but a ~ 15 -min delay was required for a second identical flash to elicit a comparable response. This light adaptation was also obvious during steady light as a relaxation of the force from a transient peak to a lower plateau level (Fig. 1d; $n = 3$ muscles). In contrast, acetylcholine-elicited contraction did not show this adaptation (Supplementary Fig. 2), suggesting that this adaptation resided in the phototransduction pathway upstream of the contractile mechanism.

The intrinsic PLR turns out to be widespread across nocturnal mammals (Supplementary Methods). Albino mice showed the same phenomenon but their sphincter muscle was more photosensitive (not shown), presumably due to higher light transmission through

¹Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ²Center for Sensory Biology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ³Department of Cardiology, Children's Hospital Boston, Harvard Medical School, and Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA. ⁴BA/MS Concurrent Program in Neuroscience, Johns Hopkins University, Baltimore, Maryland 21238, USA. ⁵Neuroscience Graduate Program, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ⁶Department of Ophthalmology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ⁷The Zanvil Krieger Mind/Brain Institute, Johns Hopkins University, Baltimore, Maryland 21238, USA. ⁸Institut für Experimentelle und Klinische Pharmakologie und Toxikologie, Universität des Saarlandes, 66421 Homburg/Saar, Germany. ⁹Division of Biology, California Institute of Technology, Pasadena, California 91125, USA. [†]Present addresses: F.M. Kirby Neurobiology Center, Department of Neurology, Children's Hospital Boston and Harvard Medical School, Boston, Massachusetts 02115, USA (M.T.H.D.); Pharmakologisches Institut Im Neuenheimer Feld 366, 69120 Heidelberg, Germany (M.F.).

Melanopsin signalling in mammalian iris and retina

T. Xue^{1,2}, M. T. H. Do^{1,2,†}, A. Riccio³, Z. Jiang^{1,2}, J. Hsieh^{1,4}, H. C. Wang^{1,5}, S. L. Merbs⁶, D. S. Welsbie⁶, T. Yoshioka^{1,7}, P. Weissgerber⁸, S. Stolz⁸, V. Flockerzi⁸, M. Freichel^{8,†}, M. I. Simon⁹, D. E. Clapham³ & K.-W. Yau^{1,2,6}

Non-mammalian vertebrates have an intrinsically photosensitive iris and thus a local pupillary light reflex (PLR). In contrast, it is thought that the PLR in mammals generally requires neuronal circuitry connecting the eye and the brain. Here we report that an intrinsic component of the PLR is in fact widespread in nocturnal and crepuscular mammals. In mouse, this intrinsic PLR requires the visual pigment melanopsin; it also requires PLC β 4, a vertebrate homologue of the *Drosophila* NorpA phospholipase C which mediates rhabdomeric phototransduction. The *Plcb4*^{-/-} genotype, in addition to removing the intrinsic PLR, also essentially eliminates the intrinsic light response of the M1 subtype of melanopsin-expressing, intrinsically photosensitive retinal ganglion cells (M1-ipRGCs), which are by far the most photosensitive ipRGC subtype and also have the largest response to light. Ablating in mouse the expression of both TRPC6 and TRPC7, members of the TRP channel superfamily, also essentially eliminated the M1-ipRGC light response but the intrinsic PLR was not affected. Thus, melanopsin signalling exists in both iris and retina, involving a PLC β 4-mediated pathway that nonetheless diverges in the two locations.

The discovery of ipRGCs has overturned the century-old belief that rods and cones are the only mammalian retinal photoreceptors^{1–6}. These ganglion-cell photoreceptors serve primarily non-image visual functions, one of which is the PLR. In animals such as fish, amphibians and bird, in addition to the neurally driven PLR, the iris itself is capable of autonomous, light-induced constriction^{7–11}. For mammals, it is thought that the PLR generally requires neuronal circuitry through the brain, although sporadic reports^{7,12,13} and some controversy exist of an intrinsic iridic photosensitivity in occasional species, including human. Even in non-mammalian vertebrates, the photopigment driving the intrinsic PLR remains unidentified. In amphibians and fish this photopigment has been proposed to be rhodopsin^{8,9}, whereas in chicken it is suggested to be the non-opsin-based cryptochrome¹¹.

We have examined the above questions more closely, and found that the intrinsic PLR is surprisingly widespread in mammals. Moreover, it bears a close relationship to the ipRGCs in phototransduction.

Intrinsic PLR in mouse and other mammals

We found that bright light triggered a pupillary constriction in an intact eye freshly isolated from a dark-adapted pigmented mouse (Fig. 1a; $n = 3$ eyes; Supplementary Methods). This photosensitivity disappeared within ~ 1 min after eye isolation, perhaps due to anoxia, with the pupil remaining constricted thereafter. This PLR persisted in a reduced preparation with just the eye's anterior chamber and iris^{8,11,12} under L-15 medium (Fig. 1b; $n = 3$ eyes; Supplementary Methods), even after blocking any potential parasympathetic activity to the iris with 0.5% atropine; however, the PLR again faded after a few light trials.

The mammalian iris has three main tissue layers, all pigmented with melanin: an anterior fibrovascular stroma; a middle smooth-muscle

layer consisting of the circumferential sphincter muscle at the pupil perimeter and the radial dilator muscle across the iris; and a posterior epithelium^{7,14}. In non-mammalian vertebrates, the sphincter muscle itself is thought to be the light sensor^{7,14}. Accordingly, we reduced the mouse preparation further to the small ring of sphincter muscle, and connected it to a strain gauge for measuring isometric tension under oxygenated Ames solution^{9,10} (35–37 °C; Supplementary Methods). The isolated sphincter muscle reproducibly gave a light-induced contractile force for hours. A relatively dim flash elicited a transient increase in force that grew linearly with increasing flash intensity; that is, proportional to flash intensity and with a constant waveform (Fig. 1c and inset, bottom; see also Supplementary Fig. 1). This flash-induced contraction reached a transient peak in ~ 1 s but decayed much more slowly, not dissimilar to the *in situ* ipRGC-driven PLR¹⁵. The response to an intense flash often showed a hump during its decay (Fig. 1c inset, top). Although the force elicited by a dim flash decayed to baseline in ≤ 1 min, a second identical flash typically elicited a smaller response unless given in ≥ 8 min after the first. Likewise, a near-saturated response to an intense flash decayed in ~ 1 min, but a ~ 15 -min delay was required for a second identical flash to elicit a comparable response. This light adaptation was also obvious during steady light as a relaxation of the force from a transient peak to a lower plateau level (Fig. 1d; $n = 3$ muscles). In contrast, acetylcholine-elicited contraction did not show this adaptation (Supplementary Fig. 2), suggesting that this adaptation resided in the phototransduction pathway upstream of the contractile mechanism.

The intrinsic PLR turns out to be widespread across nocturnal mammals (Supplementary Methods). Albino mice showed the same phenomenon but their sphincter muscle was more photosensitive (not shown), presumably due to higher light transmission through

¹Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ²Center for Sensory Biology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ³Department of Cardiology, Children's Hospital Boston, Harvard Medical School, and Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA. ⁴BA/MS Concurrent Program in Neuroscience, Johns Hopkins University, Baltimore, Maryland 21238, USA. ⁵Neuroscience Graduate Program, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ⁶Department of Ophthalmology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ⁷The Zankov Krieger Mind/Brain Institute, Johns Hopkins University, Baltimore, Maryland 21238, USA. ⁸Institut für Experimentelle und Klinische Pharmakologie und Toxikologie, Universität des Saarlandes, 66421 Homburg/Saar, Germany. ⁹Division of Biology, California Institute of Technology, Pasadena, California 91125, USA. [†]Present addresses: F.M. Kirby Neurobiology Center, Department of Neurology, Children's Hospital Boston and Harvard Medical School, Boston, Massachusetts 02115, USA (M.T.H.D.); Pharmakologisches Institut Im Neuenheimer Feld 366, 69120 Heidelberg, Germany (M.F.).

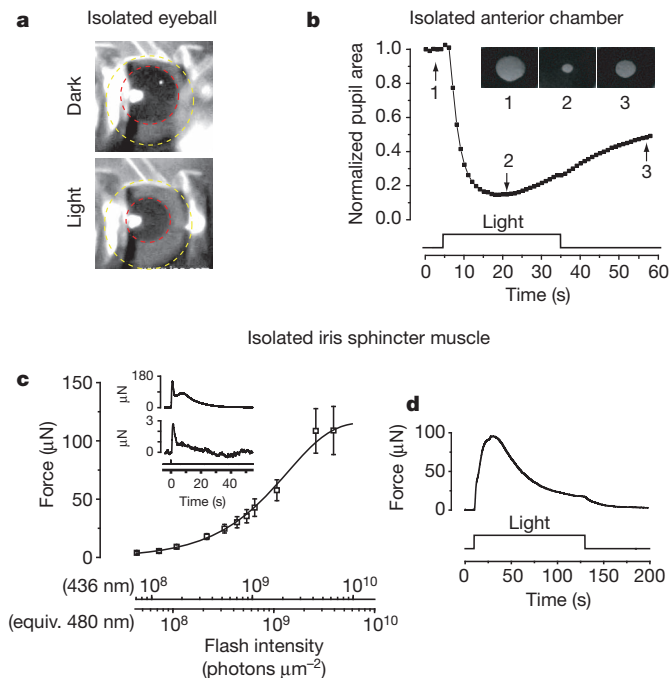


Figure 1 | Intrinsic PLR of mouse. **a**, Constriction of pupil (red circle) in freshly isolated eye (yellow circle), elicited by $1.2 \times 10^{-3} \mu\text{W} \mu\text{m}^{-2}$ of white Xe light for ~ 3 s over iris (23°C). **b**, Intrinsic PLR in isolated anterior chamber, plotted as pupil area normalized to dark state (23°C). The inset shows pupil at times indicated by arrows; white Xe light (30 s of $4 \times 10^{-4} \mu\text{W} \mu\text{m}^{-2}$) over entire iris. **c**, Flash-intensity-response relation for sphincter muscle force at transient peak of response (mean \pm s.e.m., 7 muscles) (35 – 37°C). Fit is $R_{\text{max}}(1 - e^{-I/I_0})$ with $R_{\text{max}} = 116 \mu\text{N}$, $I_0 = 2.3 \times 10^9 \text{ photons } \mu\text{m}^{-2}$ (436-nm Hg light except for brightest two intensities, which were white; 3-mm spot covering entire muscle). Flashes were 12–400 ms in duration. The inset shows sample responses to a dim and a bright flash delivering (at time 0) 7.2×10^7 and $6.5 \times 10^9 \text{ photons } \mu\text{m}^{-2}$ (436 nm), respectively. **d**, Muscle-force response to a light step ($6.1 \times 10^9 \text{ photons } \mu\text{m}^{-2} \text{ s}^{-1}$ at 436 nm) to indicate adaptation. Flash intensities are also expressed in equivalent 480-nm photons, given that melanopsin is probably the signalling pigment (see Fig. 3). All wild-type mice here and in subsequent figures were C57BL/6J, unless indicated otherwise.

the non-pigmented iris. Albino rat (Fig. 2a, left; $n = 2$ muscles; pigmented species not tested) and hamster (Fig. 2a, right; $n = 3$ muscles; pigmented, but with little iridic melanin in the pupillary margin) also tested positively, with sensitivities similar to the albino mouse but with a much larger muscle force. Dog, cat and rabbit (all pigmented) also showed an intrinsic PLR ($n = 2, 1$ and 3 muscles, respectively), but with distinctly lower photosensitivity and force production; cat and rabbit required a light step instead of flash to elicit a detectable response (Fig. 2b). Although these three species are not strictly nocturnal, they are crepuscular (that is, active at dawn and dusk). The Nile grass rat, which has both diurnal and nocturnal tendencies¹⁶, also tested positively ($n = 3$ muscles, data not shown). On the other hand, the response was absent for guinea-pig ($n = 2$), ground squirrel ($n = 4$) and pig ($n = 2$) (data not shown). Ground squirrel is strongly diurnal, whereas guinea-pig and pig are variably described as crepuscular or even diurnal. None of four primate species tested showed this phenomenon, including rhesus monkey (Fig. 2b; diurnal, $n = 7$ muscles), marmoset (diurnal, $n = 2$), owl monkey (nocturnal, $n = 4$) and bush baby (nocturnal, $n = 2$) (the latter three not shown). Thus, nocturnal/crepuscular non-primate mammals tend to have an intrinsic PLR.

Melanopsin involvement in mammalian intrinsic PLR

The action spectrum for the isolated mouse sphincter muscle (Supplementary Methods) fit an A_1 -pigment spectral template¹⁷ with λ_{max} at 480 nm (Fig. 3a), indicative of the involvement of melanopsin^{3–6}.

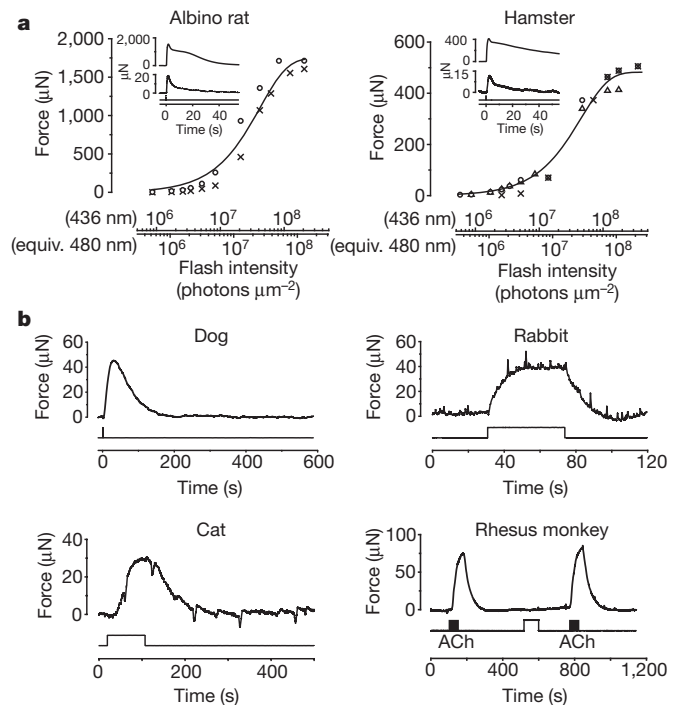


Figure 2 | Intrinsic photosensitivity of iris sphincter muscle of other mammals. Temperature range 35 – 37°C . **a**, Flash-intensity-response relations from robust light responses of albino rat and pigmented hamster muscles (two for rat and three for hamster). Same fits as in Fig. 1, with I_0 (averaged across individual muscles) being $4.3 \times 10^7 \text{ photons } \mu\text{m}^{-2}$ and $4.6 \times 10^7 \text{ photons } \mu\text{m}^{-2}$ (436 nm) for rat and hamster, respectively. Flashes were 12–400 ms. Insets show sample responses from a muscle of each species to dim and saturating flashes: 3.4×10^6 and $2.1 \times 10^8 \text{ photons } \mu\text{m}^{-2}$ for rat and 1.6×10^6 and $3.6 \times 10^8 \text{ photons } \mu\text{m}^{-2}$ for hamster, respectively (436-nm Hg light). **b**, Similar experiments on dog (representative of 2 muscles), rabbit (3 muscles), cat (1 muscle) and rhesus monkey (7 muscles). All steps and the flash (600 ms) delivered $6.1 \times 10^9 \text{ photons } \mu\text{m}^{-2} \text{ s}^{-1}$ (436 nm). Monkey muscle (pre-incubated with $30 \mu\text{M}$ 9-*cis*-retinal for 1 h) gave no obvious light response, but responded to $50 \mu\text{M}$ acetylcholine (ACh) in bath (black bars). Fast deflections in the rabbit and cat experiments reflected spontaneous muscle contractions/relaxations of unknown cause. A 5-mm-diameter light spot was used, covering the entire muscle of rat, hamster, cat and rabbit, but only partially the larger muscles of other species.

Indeed, the muscle from melanopsin knockout¹⁸ (*Opn4*^{−/−}) mice gave no light response (Fig. 3b). Using polymerase chain reaction with reverse transcription (RT-PCR), we detected melanopsin messenger RNA in the isolated wild-type mouse iris (Fig. 3c and Supplementary Methods). Immunostaining (in an albino background for viewing immunofluorescence) also indicated melanopsin's presence in the sphincter muscle of wild-type, but not *Opn4*^{−/−}, iris (Fig. 3d and Supplementary Methods). A BAC-transgenic mouse expressing the fluorescent protein tdTomato under the melanopsin promoter¹⁵ (*Opn4:tdTomato*, bred into an albino background) also showed fluorescence at least in the sphincter muscle (Fig. 3e). For rat and hamster, their particularly robust sphincter-muscle photoresponses (see above) allowed characterization of the associated action spectra, which also fit a 480-nm spectral template (Supplementary Fig. 3). Strictly speaking, because melanopsin seems to be more widespread than in just the sphincter muscle (Fig. 3d), contraction is not necessarily initiated by light absorbed just in the muscle itself, but this is likely.

For mouse, we were able to rule out any significant role of rhodopsin or cryptochromes in the sphincter-muscle photosensitivity (see above); namely, besides a null contractile phenotype of the *Opn4*^{−/−} muscle, the responses from *Rho*^{−/−} (rhodopsin knockout¹⁹) and *Cry1*^{−/−}*Cry2*^{−/−} (cryptochrome 1 and 2 double knockout²⁰) mice appeared normal (Supplementary Fig. 4a). RT-PCR detected rhodopsin mRNA in the

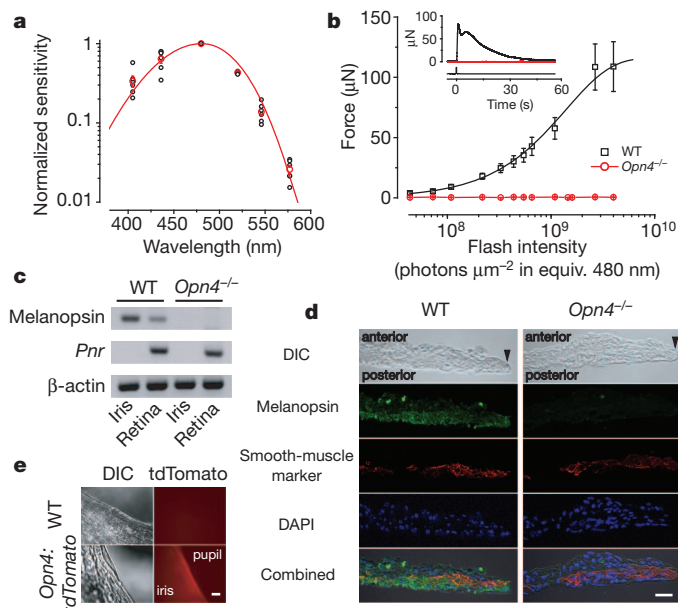


Figure 3 | Dependence of intrinsic PLR on melanopsin. **a**, Action spectrum of mouse muscle (wild type, pigmented) (6 muscles), with sensitivity normalized to value at 480 nm in each experiment. Red circles are averages, and curve is an A₁-pigment spectral template¹⁷ with $\lambda_{\text{max}} = 480$ nm. Hg light with interference filters used. Error bars are s.e.m. **b**, Average flash-intensity-response relation for *Opn4*^{-/-} muscle (3 muscles). Wild-type relation from Fig. 1c also shown for comparison. Error bars are s.e.m. The inset shows sample responses of wild-type and *Opn4*^{-/-} muscles to a saturating flash of 4.0×10^9 photons μm^{-2} (equivalent 480 nm). **c**, Melanopsin (*Opn4*) mRNA detected by RT-PCR in iris and retina from wild type but not *Opn4*^{-/-}. *Pnr* (photoreceptor-specific nuclear receptor; also called *Nr2e3*) mRNA was used as a control to rule out contamination from retina to iris. β -Actin mRNA was used as positive control. Difference in melanopsin mRNA signal between iris and retina presumably reflects different fractional total-tissue mRNA coding for melanopsin. **d**, Immunostaining of wild-type and *Opn4*^{-/-} iris cross-sections for melanopsin (green), smooth-muscle α -actin (red, as muscle marker) and DAPI (blue). Anterior side (stroma) up and posterior side (posterior epithelium) down. The black arrowhead marks the pupillary edge. The intensity of melanopsin immunofluorescence appeared lower in the iris than in ipRGCs (not shown). Additionally, although the *Opn4*^{-/-} mouse contains the *tau-lacZ* marker gene replacing *Opn4* (ref. 18), β -galactosidase activity (by X-gal labelling) was not evident in the iris (not shown), presumably due to the low melanopsin-promoter activity. Scale bar: 20 μm . **e**, TdTomato fluorescence signal detected in the iris of *Opn4:tdTomato* but not wild-type mice. Scale bar: 100 μm . Stimuli in **b** were 436-nm Hg light, and white for the two brightest intensities, although expressed in equivalent 480-nm photons. A 3-mm-diameter spot covered the entire muscle. All force measurements at 35–37 °C. Mice in **d** and **e** were albino (C57BL/6J-Tyr^{-2J}/J).

dissected mouse iris (not shown; see also ref. 21), but its significance is unclear.

By RT-PCR, we found melanopsin mRNA in the iris of rhesus monkey (not shown) and baboon (also diurnal; Supplementary Fig. 5), but the melanin pigmentation confounded confirmation by immunohistochemistry. Melanopsin's function in the primate iris is similarly unclear because there is no intrinsic PLR.

Phototransduction mechanism underlying intrinsic PLR

Melanopsin has a phylogenetic association with invertebrate rhabdomeric visual pigments²², thus possibly sharing a common phospholipase C (PLC)-mediated phototransduction pathway (see ref. 23 for review). Moreover, PLC typically mediates membrane-receptor signalling in smooth muscles^{24,25}. Indeed, the sphincter muscle from *Plcb4*^{-/-} mice²⁶ was practically unresponsive to light (Fig. 4a; $n = 5$ muscles). Sometimes, we observed a tiny response that disappeared after several stimuli (red trace in Fig. 4b, top; dim flash; 3 out of 5 muscles), unlike the much larger and persistent wild-type response (black trace in Fig. 4b, top; dim

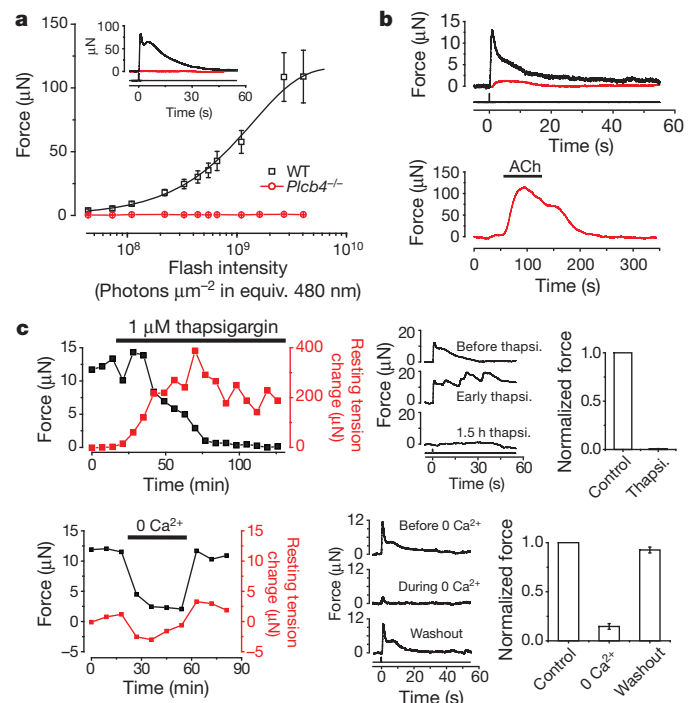


Figure 4 | Phototransduction mechanism underlying intrinsic PLR.

a, Average flash-intensity-response relation for *Plcb4*^{-/-} muscle (5 experiments). Wild-type relation from Fig. 1c shown for comparison. The inset shows sample responses of wild-type and *Plcb4*^{-/-} to a saturating flash, at 4.0×10^9 photons μm^{-2} (equivalent 480 nm). **b**, Top: example of a *Plcb4*^{-/-} muscle (red) showing a tiny response ($< 3 \mu\text{N}$) to the first few dim flashes (7.3×10^7 photons μm^{-2}) before becoming unresponsive. The wild-type response (black) to the same dim stimulus is also shown for comparison. Bottom: the same *Plcb4*^{-/-} muscle responded substantially to 10 μM acetylcholine. **c**, Thapsigargin and removal of extracellular Ca^{2+} , respectively, greatly diminished the light response. Left: time course of effect on peak force (black) generated by dim flashes (1.1×10^8 photons μm^{-2}). Resting muscle tension (red) arbitrarily set as 0 before thapsigargin or 0 Ca^{2+} application. Middle: sample responses. Right: collected data (4 muscles each). White Hg light for two brightest flashes in intensity-response relations of **a**; all other stimuli were 436-nm Hg light. All intensities are expressed in equivalent 480 nm photons. Flashes delivered at time 0, as a 3-mm-diameter spot covering the entire muscle; 35–37 °C. All error bars are s.e.m.

flash). This small response could be mediated by a different PLC- β isoform or another minor pathway. The *Plcb4*^{-/-} phenotype was not due to a defective contractile apparatus, because acetylcholine still elicited strong contraction via muscarinic receptors on the muscle¹⁴ (Fig. 4b, bottom). PLC $\beta 4$ is the closest vertebrate homologue²⁶ of the *Drosophila* PLC (NorPA), which mediates phototransduction in rhabdomeric photoreceptors²³. Bath-applied U71322, a PLC inhibitor, did not block the light-induced contraction of the wild-type sphincter muscle (not shown), but this may reflect poor drug penetration into the tissue (see below).

Smooth-muscle contraction often involves intracellular Ca^{2+} release (triggered by PLC via inositol-1,4,5-trisphosphate generation) in tandem with extracellular Ca^{2+} influx^{24,25}. Indeed, blocking intracellular Ca^{2+} uptake with 1 μM thapsigargin to deplete the Ca^{2+} -release pool gradually eliminated the muscle's light response (Fig. 4c, top; $n = 3$ muscles). The muscle's resting tension increased and oscillated during thapsigargin application, suggesting poor intracellular Ca^{2+} handling. Removing extracellular Ca^{2+} likewise reduced the light response by $\sim 80\%$ (Fig. 4c, bottom; see also refs 9, 10, 12) and partially relaxed the resting tension. Because membrane depolarization is reportedly unnecessary for the intrinsic PLR in non-mammalian vertebrates^{9,10}, Ca^{2+} -permeable ion channels other than voltage-gated Ca channels are probably involved, with TRP channels—especially

TRPC channels—being candidates²⁵. However, neither the TRPC1/4/5 nor the TRPC3/6/7 subfamily was apparently involved because the *Trpc1*^{-/-}*Trpc4*^{-/-}*Trpc5*^{-/-} and *Trpc3*^{-/-}*Trpc6*^{-/-}*Trpc7*^{-/-} genotypes appeared normal (Supplementary Methods and Supplementary Fig. 4b). A role of TRPV4 in smooth muscle has also been described²⁵, but the muscle from *Trpv4*^{-/-} mice²⁷ was also normal (Supplementary Methods and Supplementary Fig. 4b). With the apparent complexity of Ca²⁺ release and Ca²⁺ influx of unknown proportions^{10,28}, we did not attempt to dissect the mechanistic details further.

Phototransduction pathway in ipRGCs

We carried out the same interrogation on single ipRGCs (Supplementary Methods). We focused on the M1 subtype of ipRGCs, which have by far the highest photosensitivity and the largest responses^{3,4}, and are also strongly labelled in our *Opn4:tdTomato* BAC-transgenic line¹⁵. Indeed, dissociated *Plcb4*^{-/-}/*Opn4:tdTomato* M1-ipRGCs gave no detectable responses (Fig. 5a, left; *n* = 6 cells, 23 °C). In flat-mount retina, *Plcb4*^{-/-} ipRGCs (with normal impulse firing upon current injection) showed a tiny saturated response (2.7 ± 1.4 pA, *n* = 4 cells) to strong flashes ($2.1\text{--}4.0 \times 10^9$ equivalent 480-nm photons μm^{-2}), or $\leq 1\%$ of wild type (Fig. 5a, right and Fig. 5c; 23 °C). As with sphincter muscle, this small response may involve another PLC- β isoform or a minor pathway. Also, as with the iris, U71322 did not block *in situ* wild-type M1-ipRGC responses in flat-mount retina (not shown; see also ref. 29), presumably due to poor drug penetration²⁹.

Interestingly, at 23 °C for recording stability, M1-ipRGCs from *Trpc6*^{-/-}*Trpc7*^{-/-} mice (crossed into *Opn4:tdTomato* (ref. 15) for cell identification; Supplementary Methods) were not intrinsically photosensitive (Fig. 5b, c). Single-knockout *Trpc6*^{-/-} and *Trpc7*^{-/-} M1-ipRGCs had similar sensitivities and saturated photocurrents as wild type, albeit with distinct response kinetics (Fig. 5b–d). TRPC6 and TRPC7 probably form heteromeric ion channels in ipRGCs, although separate homomeric channels are remotely still possible (Supplementary Fig. 6). The *Trpc1*^{-/-}*Trpc4*^{-/-}*Trpc5*^{-/-} triple mutant and *Trpc3*^{-/-} phenotypes were both like wild type (Fig. 5b–d), whereas the *Trpc3*^{-/-}*Trpc6*^{-/-} and *Trpc3*^{-/-}*Trpc6*^{-/-}*Trpc7*^{-/-} phenotype (5 cells each, not shown) resembled *Trpc6*^{-/-} and *Trpc6*^{-/-}*Trpc7*^{-/-}, respectively. The TRPC subfamily is the closest vertebrate homologue of *Drosophila* TRP/TRPL channels which mediate rhabdomeric phototransduction downstream of NorpA (ref. 30).

At 35 °C, M1-ipRGCs (further validated by intracellular dye-labelling after recordings; Supplementary Methods) of the *Trpc6*^{-/-}*Trpc7*^{-/-} genotype did give detectable, albeit tiny, responses ($\leq 1\%$ of wild type) to strong flashes ($1.1\text{--}1.9 \times 10^9$ equivalent 480-nm photons μm^{-2}), as did *Trpc3*^{-/-}*Trpc6*^{-/-}*Trpc7*^{-/-} cells (Supplementary Fig. 7). Presumably, these residual responses would be even smaller¹⁵ and therefore not detected at 23 °C. *Opn4*^{-/-} cells, however, remained unresponsive at 35 °C (Supplementary Fig. 7).

Functional contribution of intrinsic PLR in mouse

Because of the intrinsic PLR, the overall pupil constriction in an illuminated eye should be stronger than the consensual constriction in the contralateral, unilluminated eye. Accordingly, we simultaneously monitored both pupils of a mouse while subjecting one eye to Ganzfeld illumination (2-min light step; Supplementary Methods). For wild-type animals, the intensity–response (I–R) relations for the ipsilateral and contralateral PLRs were identical at the dimmest intensities but diverged thereafter, with the ipsilateral PLR being always stronger within a given animal (Fig. 6a, b, top panel). For *Opn4*^{-/-} animals, some bilateral asymmetry persisted but it was noticeably smaller especially at high intensities, with both I–R relations being broadly similar to wild type except for a lower maximal PLR as found previously¹⁸ (Fig. 6a, b, second panel from top). For mice lacking rod and cone signals¹⁵ (*Gnat1*^{-/-}*cl*), the I–R relation was shifted to much higher light-step intensities owing to exclusive

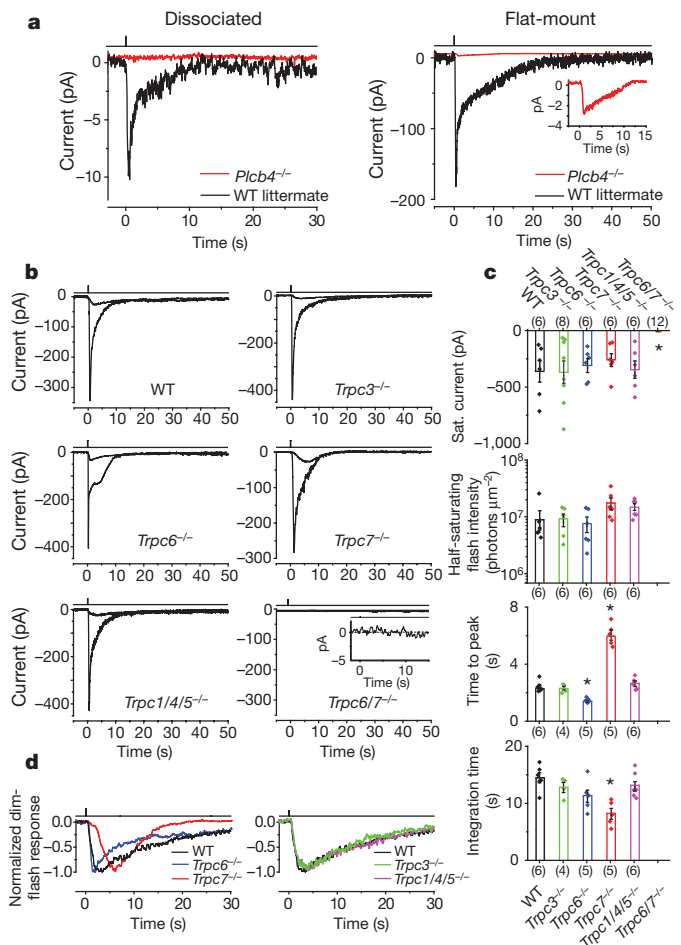


Figure 5 | Phototransduction mechanism and components underlying ipRGC intrinsic light response. Cells identified for electrical recordings based on *Opn4:tdTomato* reporter background in all mouse lines (23 °C). **a**, Left: dissociated *Plcb4*^{-/-} ipRGCs showed no detectable response to a saturating white Xe flash. An average of four trials for each response is shown. Right: in flat-mount retina after synaptic block, *Plcb4*^{-/-} ipRGCs showed a tiny residual response (magnified in inset) to similar saturating flash. Single flash trial for wild-type littermate and average of five trials for *Plcb4*^{-/-} mice, explaining the low baseline noise in *Plcb4*^{-/-}; 50–100-ms flashes delivering $1.97\text{--}3.94 \times 10^9$ photons μm^{-2} (equivalent 480 nm) in all cases. **b**, Intrinsic light responses from ipRGCs in flat-mount retina from wild-type, *Trpc3*^{-/-}, *Trpc6*^{-/-}, *Trpc7*^{-/-}, *Trpc1*^{-/-}*Trpc4*^{-/-}*Trpc5*^{-/-} and *Trpc6*^{-/-}*Trpc7*^{-/-} mice. Sample responses to both dim ($2.28\text{--}7.76 \times 10^5$ photons μm^{-2}) and saturating flashes (2.05×10^8 photons μm^{-2}) in each case are shown except for *Trpc6*^{-/-}*Trpc7*^{-/-}, for which only response to brightest flash is shown (magnified in inset and essentially zero). Flashes were 100–300 ms of 505-nm LED light. **c**, Comparison of saturated response amplitude and dim-flash-response parameters from **b**. Half-saturating flash intensity (that is, intensity eliciting a half-maximum response) is inversely proportional to sensitivity. Time-to-peak is the time lapse between flash and transient peak of dim-flash response. Integration time, t_i , of the dim-flash response is a measure of its overall duration, given by $t_i = \int f(t)dt/f_p$, where $f(t)$ is response profile and f_p is its transient-peak amplitude. Collected data (mean \pm s.e.m.), where asterisk indicates $P < 0.001$ compared to wild type by two-sample *t*-test. The number of cells examined for each parameter is in parentheses, and not always the same across parameters because not all parameters were obtainable for each cell. **d**, Normalized, averaged dim-flash responses from **b** for comparison of response waveforms. All intensities expressed in equivalent 480-nm photons, and delivered as a 730- μm -diameter spot centred on soma, sufficient for covering entire intact cell in retina. Light monitor above each trace. All recordings were in perforated-patch, voltage-clamp mode ($V_{\text{hold}} = -80$ mV), with synaptic transmission blocked pharmacologically.

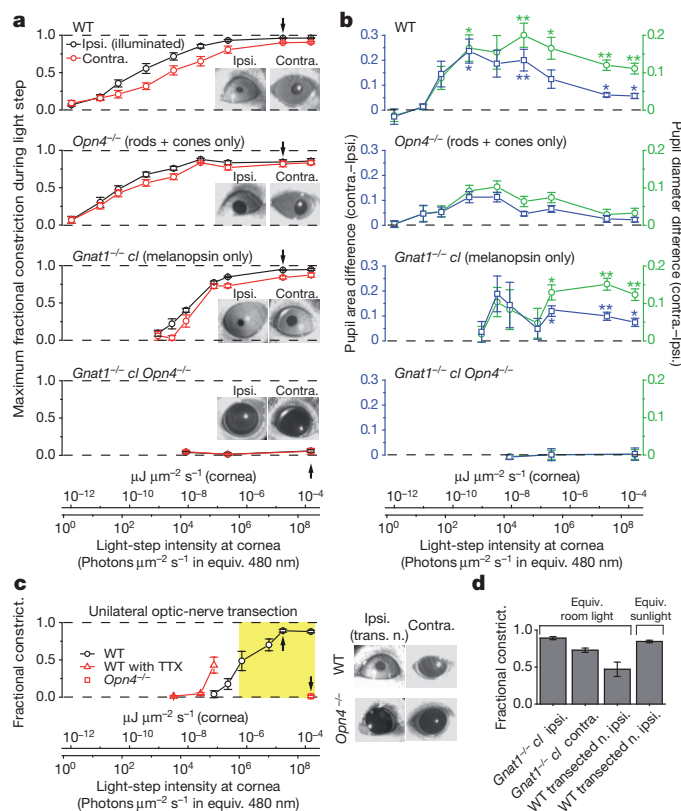


Figure 6 | Simultaneous direct (ipsilateral) and consensual (contralateral) PLRs to unilateral illumination for different mouse genotypes *in situ*. Illumination for 2 min with 505-nm LED light in **a–c**. PLR in ipsilateral, illuminated eye was measured at peak during this 2-min period, with contralateral PLR measured simultaneously. For **a** and **b**, conversion of light intensities into equivalent 480-nm photons applies strictly only to *Gnat1*^{-/-} cl but not to wild-type or *Opn4*^{-/-} genotypes (which involved also rod/cone signals), because it required action spectrum of melanopsin. For this reason, light intensities are also given in the general unit of $\mu\text{J} \mu\text{m}^{-2} \text{s}^{-1}$ (same as $\mu\text{W} \mu\text{m}^{-2}$). **a**, Average step-intensity-response (I–R) relations for wild-type, *Opn4*^{-/-}, *Gnat1*^{-/-} cl and *Gnat1*^{-/-} cl *Opn4*^{-/-} mice. PLR is expressed as maximum fractional constriction (MFC), where MFC = 1 – (normalized pupil area in light) = 1 – (pupil area in light/pupil area in darkness). The inset shows exemplary bilateral PLRs at an intensity indicated by an arrow on I–R relations. Number of animals: 7 (wild type), 7 (*Opn4*^{-/-}), 5 (*Gnat1*^{-/-} cl) and 3 (*Opn4*^{-/-} *Gnat1*^{-/-} cl). **b**, Ipsi/contralateral difference in normalized pupil area and diameter. For area, this difference is simply the difference between the ipsilateral and contralateral values in **a**. For diameter, normalized diameter is given by (normalized area)^{1/2} = (1 – MFC)^{1/2}, calculable from data in **a**; the ipsi/contralateral difference values can then be evaluated (see Supplementary Methods for significance of diameter). The area or diameter difference at a given intensity was then averaged over all animals of a given genotypic group. **P* < 0.05, ***P* < 0.01, when a wild-type or *Gnat1*^{-/-} cl value is compared to corresponding *Opn4*^{-/-} value by a two-sample *t*-test. **c**, Direct (ipsilateral) PLR of eye with transected optic nerve to isolate the intrinsic component. Data for wild-type mice in the absence or presence of TTX (5 μl of 600 μM TTX in water administered on the cornea) (same set of five animals), and *Opn4*^{-/-} mice (4 animals) are shown. Images on right show exemplary bilateral PLRs (without TTX) at an intensity indicated by arrow on corresponding I–R relations. *Opn4*^{-/-} animals with transected optic nerve gave essentially no PLR. Yellow shaded area indicates light-intensity range from room light (with a minimum measured as $1.0 \times 10^{-6} \mu\text{W} \mu\text{m}^{-2}$, or 1.1×10^6 equivalent 480-nm photons $\mu\text{m}^{-2} \text{s}^{-1}$) to direct sunlight (measured as $3.7 \times 10^{-4} \mu\text{W} \mu\text{m}^{-2}$, or 3.9×10^8 equivalent 480-nm photons $\mu\text{m}^{-2} \text{s}^{-1}$) (Supplementary Methods). **d**, Pupil constriction triggered by white light (400–650-nm band-pass filter; Xe lamp), matched in power to average common room light and ambient sunlight (Supplementary Methods) (2.2×10^{-6} and $4.2 \times 10^{-5} \mu\text{W} \mu\text{m}^{-2}$, respectively), for *Gnat1*^{-/-} cl without optic-nerve transection (6 animals) and wild-type mice with optic-nerve transection (5 animals). All light was rendered adirectional with a Ganzfeld diffusing sphere. Pupils were video-recorded under dim infrared light that did not activate any photoreceptors. All error bars are s.e.m.

signalling by melanopsin¹⁸, but the bilateral asymmetry at high intensities was also more obvious than for *Opn4*^{-/-} mice (Fig. 6a, b, third panel from top). The residual asymmetry in *Opn4*^{-/-} may also indicate slightly stronger rod/cone signals to the ipsilateral PLR, perhaps explaining some of the wild-type ipsi/contralateral disparity in PLR especially at lower intensities. Finally, *Gnat1*^{-/-} cl *Opn4*^{-/-} mice had practically no steady PLR (ref. 31, but see ref. 32) (Fig. 6a, b, bottom panel).

The difference in bilateral asymmetry between wild-type and *Opn4*^{-/-} PLRs detailed above cannot distinguish between an intrinsic iridic PLR and a bilateral asymmetry in ipRGC signalling to the PLR because both mechanisms involve melanopsin. To isolate the intrinsic PLR, we eliminated retinal signalling from one eye in the wild-type mouse by transecting its optic nerve (Supplementary Methods). When the denervated eye was illuminated (at >7 days after surgery), the intact contralateral eye failed to respond as expected, whereas the PLR persisted in the denervated eye (Fig. 6c, right panel) with the action spectrum of melanopsin (Supplementary Fig. 8). This residual component is the isolated intrinsic PLR. Its I–R relation on the intensity axis (Fig. 6c, left panel) relative to that for the ipsilateral PLR of non-operated wild-type animals (Fig. 6a, top panel) indicates that the intrinsic PLR begins to contribute when the normal overall PLR is ~90% complete. Nonetheless, the intrinsic PLR even by itself would have been able to drive the pupillary constriction 80–90% to completion over ~3 log units of light-step intensities. Furthermore, the intrinsic PLR has, in reality, an even lower light threshold (thus contributing even more to the overall PLR) because its I–R relation shifted by ~1 log unit to lower intensities after topical application of tetrodotoxin (TTX) to the cornea (Fig. 6c, left panel), which blocked any tonic autonomic inputs to the dilator and sphincter muscles (Supplementary Methods). In short, the intrinsic component participates in the highest ~4 log units of the overall ~9-log-unit dynamic range of light intensities spanned by the normal PLR in mouse, at least during 2 min of steady illumination.

To translate into natural conditions, the yellow-shaded region in Fig. 6c, left panel, spans approximately from laboratory light to outdoor daylight (Supplementary Methods). We also directly simulated ambient light with white xenon-arc light (400–650 nm) of matched power (Supplementary Methods). For example, in room light, an ipsilateral (that is, intrinsic) PLR of 0.47 ± 0.10 fractional constriction was elicited from the denervated eye of wild-type mice (Fig. 6d, *n* = 5 animals), versus 0.89 ± 0.02 when ipRGCs were also active (*Gnat1*^{-/-} cl mice without transected optic nerve) (Fig. 6d, *n* = 6 animals). Thus, the intrinsic PLR contributes substantially to the melanopsin component of the overall PLR even in room light. Finally, exposing a dark-adapted mouse with bilaterally denervated eyes to genuine room light or outdoor light led to intrinsic PLRs to the extent expected from above (data not shown).

Similar denervation experiments on rhesus and owl monkeys revealed no intrinsic PLR (*n* = 2 animals each; Supplementary Text), consistent with the above negative findings from their sphincter muscles and with clinical observations from human patients presenting complete unilateral optic neuropathy (resulting in loss of photosensitivity in the affected eye) (Supplementary Text).

Discussion

We have discovered a surprising intrinsic PLR widespread among non-primate mammals, with a strong positive correlation between a nocturnal/crepuscular habitat and an intrinsic PLR. An intrinsic PLR benefits such mammals (with highly photosensitive, rod-dominant retinas more susceptible to photodamage³³) by sustaining a pupil constriction under strong steady light—a situation otherwise difficult to achieve with neural circuitry alone because of the meagre light admitted through the tiny pupil to the retina. The reason for the loss of this feature in primates despite melanopsin's continued presence in the iris is unclear. Additionally, in contrast to previous reports^{34,35}, our

work reveals some bilateral asymmetry in the PLR of normal mouse, which we now know to come partly from the intrinsic iridic photosensitivity in the higher-intensity range.

The intrinsic PLR is driven by melanopsin in mouse, and probably also rat and hamster. By extension, the same presumably applies to other mammals showing an intrinsic PLR. As such, the function, if any, of the rhodopsin in mouse iris (see above) is unknown. For amphibians and fish, the evidence for rhodopsin driving the intrinsic PLR was suggestive but indefinite^{8,9} and probably should be re-examined given our present findings and melanopsin's presence in the eyes of amphibians (retina and iris)²² and fish (ref. 36; H.-W. Liao and K.-W.Y., unpublished). In chicken, the intrinsic PLR reportedly has an action spectrum with λ_{max} at <400 nm (ref. 11; confirmed by us but not shown). This is interesting because chicken also has melanopsin in the retina and iris^{11,37}.

Because melanopsin is phylogenetically related to rhabdomeric opsins²², it is perhaps not surprising that its photosignalling mechanism uses a PLC pathway. Previously, clues from heterologous expression^{38,39}, pharmacology, electrophysiology and/or immunohistochemistry on frog melanophores⁴⁰, native ipRGCs²⁹ and subvertebrate chordates^{41,42} have all implicated such a pathway. With gene-knockout mouse lines, we have now established this pathway more definitively and identified the signalling enzyme as PLC β 4, the closest homologue of *Drosophila* NorpA and the PLC that mediates rhabdomeric phototransduction in fly. Although PLC β 4 is shared by the iris and ipRGCs (at least the M1 subtype), the pathway diverges downstream. For M1-ipRGCs, the depolarizing light response results from the opening of presumably heteromeric channels formed predominantly by TRPC6 and TRPC7 (with potentially additional non-TRPC or non-critical TRPC subunits), gated possibly by phosphatidylinositol-4,5-bisphosphate (ref. 29). This composition of the native channel agrees with indirect suggestions from immunohistochemical^{43,44} and RT-PCR⁴⁵ studies. For the sphincter muscle, however, TRPC3/6/7 or TRPC1/4/5 are apparently not involved in the light-induced contraction. An intracellular Ca^{2+} release is clearly important for the intrinsic PLR (see above), but apparently not for the ipRGC light response²⁹.

We have not yet identified the $G\alpha$ species signalling between melanopsin and PLC β 4. However, conventional wisdom³⁰, pharmacological evidence²⁹ and *in vitro* biochemistry⁴⁶ suggest that it (they) should belong to the G_q subfamily.

Finally, ordinary retinal ganglion cells (that is, non-ipRGCs) become intrinsically photosensitive when transduced by virus to express melanopsin⁴⁷. Because PLC β 4 and TRPC6 are both expressed generally in RGCs^{29,43,44,48}, the same signalling pathway may well underlie this virus-induced intrinsic photosensitivity.

Note added in proof: After we had completed the experiments on ipRGCs, a paper was published⁴⁹ reporting that the *Trpc3*^{-/-} and *Trpc7*^{-/-} single-knockout genotypes showed no effect on the light response of ipRGCs, whereas the *Trpc6*^{-/-} genotype showed a smaller saturated light response than wild type. These data are not in quantitative agreement with those reported here. We attribute this difference to the use of whole-cell recording by the other study, a recording configuration that, from our experience, does not monitor the light response of ipRGCs with the same fidelity as the perforated-patch recording adopted by us.

METHODS SUMMARY

Except for *Trpc5*^{-/-} and *Trpc7*^{-/-}, other single-knockout mouse lines in this study were published previously. Pupil responses of the isolated eye and the isolated anterior chamber were tested at 23 °C (room temperature). Isometric force of the isolated iridic sphincter muscle was measured at 35–37 °C with a force transducer. *In vivo* pupil constriction was measured as previously described¹⁵, but with added video monitoring of the illuminated eye, and the mice were head-anchored⁵⁰. The optic nerve of an *in vivo* eye was transected without disturbing the ocular blood supply. For electrophysiology, flat-mounted whole retina and dissociated ipRGCs were prepared as previously described¹⁵; dim excitation

flashes were used to minimize photobleaching, thus identifying only ipRGCs with especially bright tdTomato fluorescence. Such cells should be M1-ipRGCs, which in wild-type and some knockout genotypes could be validated by their high relative photosensitivity and very large photocurrents. In later experiments at ~35 °C, additional validation of the cells as the M1-subtype was made by post-recording intracellular dye-labelling and ascertaining their dendritic arborizations to be in the off-sublamina of the inner plexiform layer^{3,4}. All recordings were in the perforated-patch, voltage-clamp mode at $V_{\text{hold}} = -80$ mV, with junction potential corrected and series resistance monitored¹⁵. Synaptic transmission was blocked pharmacologically in recordings from the whole retina¹⁵. Light stimuli were from a Xe or Hg arc lamp or light-emitting diodes (LEDs). Immunohistochemistry, biochemistry and PCR experiments followed standard protocols. All indicated errors are standard errors of the mean (s.e.m.). All experimental procedures on animals followed the guidelines of the Animal Care and Use Committee of the Johns Hopkins University School of Medicine.

Received 11 May; accepted 15 September 2011.

- Berson, D. M., Dunn, F. A. & Takao, M. Phototransduction by retinal ganglion cells that set the circadian clock. *Science* **295**, 1070–1073 (2002).
- Hattar, S., Liao, H.-W., Takao, M., Berson, D. M. & Yau, K.-W. Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science* **295**, 1065–1070 (2002).
- Bailes, H. J. & Lucas, R. J. Melanopsin and inner retinal photoreception. *Cell. Mol. Life Sci.* **67**, 99–111 (2010).
- Do, M. T. H. & Yau, K.-W. Intrinsically photosensitive retinal ganglion cells. *Physiol. Rev.* **90**, 1547–1581 (2010).
- Hankins, M. W., Peirson, S. N. & Foster, R. G. Melanopsin: an exciting photopigment. *Trends Neurosci.* **31**, 27–36 (2008).
- Nayak, S. K., Jegla, T. & Panda, S. Role of a novel photopigment, melanopsin, in behavioral adaptation to light. *Cell. Mol. Life Sci.* **64**, 144–154 (2007).
- Barr, L. Photomechanical coupling in the vertebrate sphincter papillae. *Crit. Rev. Neurobiol.* **4**, 325–366 (1989).
- Seliger, H. H. Direct action of light in naturally pigmented muscle fibers. I. Action spectrum for contraction in eel iris sphincter. *J. Gen. Physiol.* **46**, 333–342 (1962).
- Barr, L. & Alpern, M. Photosensitivity of the frog iris. *J. Gen. Physiol.* **46**, 1249–1265 (1963).
- Kargacin, G. J. & Detwiler, P. B. Light-evoked contraction of the photosensitive iris of the frog. *J. Neurosci.* **5**, 3081–3087 (1985).
- Tu, D. C., Batten, M. L., Palzewski, K. & Van Gelder, R. N. Nonvisual photoreception in the chick iris. *Science* **306**, 129–131 (2004).
- Bito, L. Z. & Turansky, D. G. Photoactivation of pupillary constriction in the isolated *in vitro* iris of a mammal (*Mesocricetus auratus*). *Comp. Biochem. Physiol.* **50A**, 407–413 (1975).
- Lau, K. C., So, K. F., Campbell, G. & Lieberman, A. R. Pupillary constriction in response to light in rodents, which does not depend on central neural pathways. *J. Neurosci.* **13**, 70–79 (1992).
- Oyster, C. W. *The Human Eye: Structure and Function* (Sinauer, 1999).
- Do, M. T. H. et al. Photon capture and signalling by melanopsin retinal ganglion cells. *Nature* **457**, 281–287 (2009).
- Blanchong, J. A., McElhinny, T. L., Mahoney, M. M. & Smale, L. Nocturnal and diurnal rhythms in the unstriped Nile rat, *Arvicanthis niloticus*. *J. Biol. Rhythms* **14**, 364–377 (1999).
- Govardovskii, V. I., Fyhrquist, N., Reuter, T., Kuzmin, D. G. & Donner, K. In search of the visual pigment template. *Vis. Neurosci.* **17**, 509–528 (2000).
- Lucas, R. J. et al. Diminished pupillary light reflex at high irradiances in melanopsin-knockout mice. *Science* **299**, 245–247 (2003).
- Lem, J. et al. Morphological, physiological, and biochemical changes in rhodopsin knockout mice. *Proc. Natl Acad. Sci. USA* **96**, 736–741 (1999).
- Vitaterna, M. H. et al. Differential regulation of mammalian period genes and circadian rhythmicity by cryptochromes 1 and 2. *Proc. Natl Acad. Sci. USA* **96**, 12114–12119 (1999).
- Ghosh, S., Salvador-Silva, M. & Coca-Prados, M. The bovine iris-ciliary epithelium expresses components of rod phototransduction. *Neurosci. Lett.* **370**, 7–12 (2004).
- Provencio, I., Jiang, G., De Grip, W. J., Hayes, W. P. & Rollag, M. D. Melanopsin: An opsin in melanophores, brain, and eye. *Proc. Natl Acad. Sci. USA* **95**, 340–345 (1998).
- Yau, K.-W. & Hardie, R. C. Phototransduction motifs and variations. *Cell* **139**, 246–264 (2009).
- Berridge, M. J. Smooth muscle cell calcium activation mechanisms. *J. Physiol.* **586**, 5047–5061 (2008).
- Gonzalez-Cobos, J. C. & Trebak, M. TRPC channels in smooth muscle cells. *Front. Biosci.* **15**, 1023–1039 (2010).
- Jiang, H. et al. Phospholipase C β 4 is involved in modulating the visual response in mice. *Proc. Natl Acad. Sci. USA* **93**, 14598–14601 (1996).
- Suzuki, M., Mizuno, A., Kodaira, K. & Imai, M. Impaired pressure sensation in mice lacking TRPV4. *J. Biol. Chem.* **278**, 22664–22668 (2003).
- Zucker, R. & Nolte, J. Light-induced calcium release in a photosensitive vertebrate smooth muscle. *Nature* **274**, 78–80 (1978).
- Graham, D. M. et al. Melanopsin ganglion cells use a membrane-associated rhabdomeric phototransduction cascade. *J. Neurophysiol.* **99**, 2522–2532 (2008).
- Venkatachalam, K. & Montell, C. TRP channels. *Annu. Rev. Biochem.* **76**, 387–417 (2007).

31. Hattar, S. *et al.* Melanopsin and rod-cone photoreceptive systems account for all major accessory visual functions in mice. *Nature* **424**, 75–81 (2003).
32. Allen, A. E., Cameron, M. A., Brown, T. M., Vugler, A. A. & Lucas, R. J. Visual responses in mice lacking critical components of all known retinal phototransduction cascades. *PLoS ONE* **5**, e15063 (2010).
33. Organisciak, D. T. & Vaughan, D. K. Retinal light damage: Mechanisms and protection. *Prog. Retin. Eye Res.* **29**, 113–134 (2010).
34. Grozdanic, S. *et al.* Characterization of the pupil light reflex, electroretinogram and tonometric parameters in healthy mouse eyes. *Curr. Eye Res.* **26**, 371–378 (2003).
35. Zhu, Y. *et al.* Melanopsin-dependent persistence and photopotential of murine pupillary light responses. *Invest. Ophthalmol. Vis. Sci.* **48**, 1268–1275 (2007).
36. Bellingham, J., Whitmore, D., Philp, A. R., Wells, D. J. & Foster, R. G. Zebrafish melanopsin: isolation, tissue localisation and phylogenetic position. *Brain Res. Mol. Brain Res.* **107**, 128–136 (2002).
37. Chaurasia, S. S. *et al.* Molecular cloning, localization and circadian expression of chicken melanopsin (Opn4): differential regulation of expression in pineal and retinal cell types. *J. Neurochem.* **92**, 158–170 (2005).
38. Panda, S. *et al.* Illumination of the melanopsin signaling pathway. *Science* **307**, 600–604 (2005).
39. Qiu, X. *et al.* Induction of photosensitivity by heterologous expression of melanopsin. *Nature* **433**, 745–749 (2005).
40. Isoldi, M. C., Rollag, M. D., Castrucci, A. M. & Provencio, I. Rhabdomeric phototransduction initiated by the vertebrate photopigment melanopsin. *Proc. Natl Acad. Sci. USA* **102**, 1217–1221 (2005).
41. Gomez, M., del P., Angueyra, J. M. & Nasi, E. Light-transduction in melanopsin-expressing photoreceptors of Amphioxus. *Proc. Natl Acad. Sci. USA* **106**, 9081–9086 (2009).
42. Koyanagi, M., Kubokawa, K., Tsukamoto, H., Shichida, Y. & Terakita, A. Cephalochordate melanopsin: evolutionary linkage between invertebrate visual cells and vertebrate photosensitive retinal ganglion cells. *Curr. Biol.* **15**, 1065–1069 (2005).
43. Warren, E. J., Allen, C. N., Brown, R. L. & Robinson, D. W. The light-activated signaling pathway in SCN-projecting rat retinal ganglion cells. *Eur. J. Neurosci.* **23**, 2477–2487 (2006).
44. Sekaran, S. *et al.* 2-Aminoethoxydiphenylborane is an acute inhibitor of directly photosensitive retinal ganglion cell activity *in vitro* and *in vivo*. *J. Neurosci.* **27**, 3981–3986 (2007).
45. Hartwick, A. T. *et al.* Light-evoked calcium responses of isolated melanopsin-expressing retinal ganglion cells. *J. Neurosci.* **27**, 13468–13480 (2007).
46. Terakita, A. *et al.* Expression and comparative characterization of Gq-coupled invertebrate visual pigments and melanopsin. *J. Neurochem.* **105**, 883–890 (2008).
47. Lin, B., Koizumi, A., Tanaka, N., Panda, S. & Masland, R. H. Restoration of visual function in retinal degeneration mice by ectopic expression of melanopsin. *Proc. Natl Acad. Sci. USA* **105**, 16009–16014 (2008).
48. Peng, Y. W. *et al.* Identification of components of a phosphoinositide signaling pathway in retinal rod outer segments. *Proc. Natl Acad. Sci. USA* **94**, 1995–2000 (1997).
49. Perez-Leighton, C. E., Schmidt, T. M., Abramowitz, J., Birnbaumer, L. & Kofuji, P. Intrinsic phototransduction persists in melanopsin-expressing ganglion cells lacking diacylglycerol-sensitive TRPC subunits. *Eur. J. Neurosci.* **33**, 856–867 (2011).
50. Cahill, H. & Nathans, J. The optokinetic reflex as a tool for quantitative analyses of nervous system function in mice: application to genetic and drug-induced variation. *PLoS ONE* **3**, e2055 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the following individuals for providing knockout mouse lines: M. Caterina (*Trpv4*^{−/−}), J. Lem (*Rho*^{−/−} and *Gnat1*^{−/−}), J. C. Chen (*Rho*^{−/−}), J. Nathans (*cl*, also known as *cone-DTA*), A. Sancar (*Cry1*^{−/−} *Cry2*^{−/−}) and L. Birnbaumer (*Trpc1*^{−/−}, *Trpc3*^{−/−} and *Trpc6*^{−/−}). We thank D. Marshak, R. von der Heydt, X. Wang and V. Casagrande for eyes from baboon, rhesus monkey, marmoset and bush baby, respectively, and W. Li, R. Mi, B. O'Rourke, L. Pipitone, D. Ruben, D. Ryugo, L. Smale and G. Tomaselli for eyes of other animals. Experiments on bush baby were carried out in the Casagrande laboratory with help and hospitality. We also thank W. Gao for suggestions on the force transducer, F. Rieke and A. Sampath for suggestions on the design of the LED light system, P. Ala-Laurila for the method of equivalent 480-nm-photon conversion, H. Cahill for the mouse-head-anchoring method, X. Ren for help on western blots, O. Garalde and A. Chen for help in the monkey experiments, and W. W. S. Yue for help on RT-PCR. We also thank T. Shelley for fabricating all custom equipment, S. Kulason for help in data analysis, and L. Ding for mouse-genotyping support. Members of the Yau laboratory provided comments on the manuscript. This work was supported by US NIH Grant EY14596 and the António Champalimaud Vision Award (Portugal) to K.-W.Y.

Author Contributions T.X., M.T.H.D. and K.-W.Y. designed the experiments. T.X. and M.T.H.D. did the isolated-eyeball experiments. T.X. carried out all muscle recordings and also *Trpc* knockout ipRGC recordings at 23 °C (recordings at 35 °C were by Z.J.), the mouse-optic-nerve transections, some of the *in situ* PLR measurements (the rest done by J.H.), some of the RT-PCR experiments on melanopsin (the rest done by H.C.W.), the tdTomato-fluorescence experiments, and the generation and maintenance of all double, triple and quadruple genetically engineered mouse lines for this study. The recordings from ipRGCs of *Plcb4*^{−/−} mice and their wild-type littermates were done by M.T.H.D. (23 °C) and Z.J. (35 °C). T.X., M.T.H.D. and J.H. designed the head-fixed, dual-PLR-recording instrument. T.X. and H.C.W. did the anterior-chamber measurements. H.C.W. did the immunocytochemistry and X-gal labelling. S.L.M. and D.S.W. did the surgery of transecting the optic nerve in anaesthetized monkeys, and tested the PLR with T.X. and M.T.H.D. T.Y. participated in the monkey experiments. A.R. in D.E.C.'s laboratory made the *Trpc7*^{−/−} mouse line and did the associated characterization (retinal RT-PCR done by T.X.). P.W., S.S., V.F. and M.F. made the *Trpc5*^{−/−} line and did the associated characterization, and also provided the *Trpc4*^{−/−} and *Trpc4*^{−/−} *Trpc5*^{−/−} lines. M.I.S. provided the *Plcb4*^{−/−} line. T.X., M.T.H.D. and Z.J. analysed the data with assistance from J.H., and, together with K.-W.Y., wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.X. (txue77@gmail.com) or K.-W.Y. (kwyau@mail.jhmi.edu).

CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing

Sanjeev Shukla¹, Ersen Kavak^{2,3}, Melissa Gregory¹, Masahiko Imashimizu⁴, Bojan Shutinoski¹, Mikhail Kashlev⁴, Philipp Oberdoerffer¹, Rickard Sandberg^{2,3} & Shalini Oberdoerffer¹

Alternative splicing of pre-messenger RNA is a key feature of transcriptome expansion in eukaryotic cells, yet its regulation is poorly understood. Spliceosome assembly occurs co-transcriptionally, raising the possibility that DNA structure may directly influence alternative splicing. Supporting such an association, recent reports have identified distinct histone methylation patterns, elevated nucleosome occupancy and enriched DNA methylation at exons relative to introns. Moreover, the rate of transcription elongation has been linked to alternative splicing. Here we provide the first evidence that a DNA-binding protein, CTCF-binding factor (CTCF), can promote inclusion of weak upstream exons by mediating local RNA polymerase II pausing both in a mammalian model system for alternative splicing, CD45, and genome-wide. We further show that CTCF binding to CD45 exon 5 is inhibited by DNA methylation, leading to reciprocal effects on exon 5 inclusion. These findings provide a mechanistic basis for developmental regulation of splicing outcome through heritable epigenetic marks.

It is estimated that greater than 90% of human genes undergo alternative splicing of pre-mRNA^{1,2} and aberrant splicing has been implicated in a number of human diseases³. Alternative splicing decisions are determined by the ability of weak splice sites to effectively compete with strong splice sites for detection by the spliceosome⁴. The balance between splice site selection is principally influenced by two variables⁵: (1) the availability of splicing factors that detect enhancer or silencer sequences encoded within nascent RNA^{4,6} and (2) the rate of RNA polymerase II (pol II) transcription elongation, wherein a slow rate favours co-transcriptional spliceosome assembly at weak splice sites^{7,8}.

A surprising result of genome-wide chromatin-immunoprecipitation-sequencing (ChIP-seq) studies is the non-random distribution of several epigenetic marks in exons relative to introns. In particular, exons show elevated nucleosome density, DNA methylation of cytosine, and overrepresentation of certain histone modifications, relative to introns^{9–15}. Differential ‘marking’ of exons on DNA highlighted a possible connection between DNA structure and co-transcriptional RNA processing. Accordingly, several recent studies suggest that exonic histone modification may affect variable inclusion of alternative exons^{16,17}. Collectively, these studies raise the intriguing possibility that epigenetic modifications are maintained on DNA to aid the spliceosome in the process of exon definition^{9,13}, and that differential chromatin assembly may represent a critical aspect of alternative splicing regulation.

Processing of CD45 pre-mRNA (also known as *PTPRC*) is a well established model system to study the regulatory mechanisms of alternative splicing. CD45 is a trans-membrane protein tyrosine phosphatase that initiates signalling through antigen receptors by dephosphorylating the inhibitory tyrosine on Src family kinases¹⁸. Variable exclusion of exons 4–6 (A–C) of CD45 transcripts is tightly correlated with stages of lymphocyte development and expressed splice variants can be distinguished using isoform-specific antibodies and flow cytometry¹⁹ (Supplementary Fig. 1). In general, the larger, exon 4-containing isoforms (CD45RA) are expressed early in peripheral lymphocyte

development, whereas the shortest isoform (CD45RO), which lacks all three variable exons, is expressed in terminally differentiated lymphocytes¹⁹. We recently identified heterogeneous ribonucleoprotein L-like (hnRNPLL) as a tissue-specific master regulator of the CD45RA to CD45RO transition in peripheral lymphocytes²⁰. hnRNPLL binds to exons 4 and 6 of CD45 mRNA and blocks the inclusion of both exons in the mature message²¹. In contrast, hnRNPLL expression does not influence exclusion of exon 5 (refs 20, 22; Supplementary Fig. 2a). *In vitro* studies aimed at uncovering regulators of exon 5 exclusion have identified several ubiquitously expressed splicing factors^{22,23}. However, peripheral lymphocytes retain exon 5 until the terminal stages of development²⁴ (Supplementary Fig. 2b), portending a yet uncovered layer of regulation.

CTCF regulates exon 5 inclusion in CD45 mRNA

Considering the growing evidence for DNA-mediated regulation of spliceosome assembly, we explored the hypothesis that exon 5 inclusion is mediated by the epigenetic structure of the gene encoding CD45. By analysing published ChIP-seq data within the UCSC genome browser^{25,26}, we identified a strong CTCF peak overlapping with exon 5 across cell types. CTCF is an 11 zinc-finger DNA-binding protein with multiple nuclear functions, largely grouped into two categories: insulating inactive regions of the genome from active regions and promoting long-range interactions between distal regions of the genome^{27,28}. Whereas intergenic CTCF is an effective barrier to transcription²⁷, we found that CTCF binding at exon 5 is maintained in cells that actively transcribe abundant CD45 (ref. 26) and that exon 5 binding is conserved in murine splenocytes (Fig. 1a and Supplementary Fig. 2c), indicating an important, position-dependent ‘non-insulator’ function. We thus explored whether and how CTCF binding at CD45 exon 5 DNA influences processing of CD45 transcripts.

To dissect the impact of CTCF on exon 5 splicing in a cell-based system, we screened several human Burkitt lymphoma B cell lines for differences in expression of the exon 5 containing CD45RB isoform.

¹Center for Cancer Research, Mouse Cancer Genetics Program, National Cancer Institute at Frederick, Frederick, Maryland 21702, USA. ²Department of Cell and Molecular Biology, Karolinska Institutet, SE-171 77 Stockholm, Sweden. ³Ludwig Institute for Cancer Research, SE-171 77, Stockholm, Sweden. ⁴Center for Cancer Research, Gene Regulation and Chromosome Biology Laboratory, National Cancer Institute at Frederick, Frederick, Maryland 21702, USA.

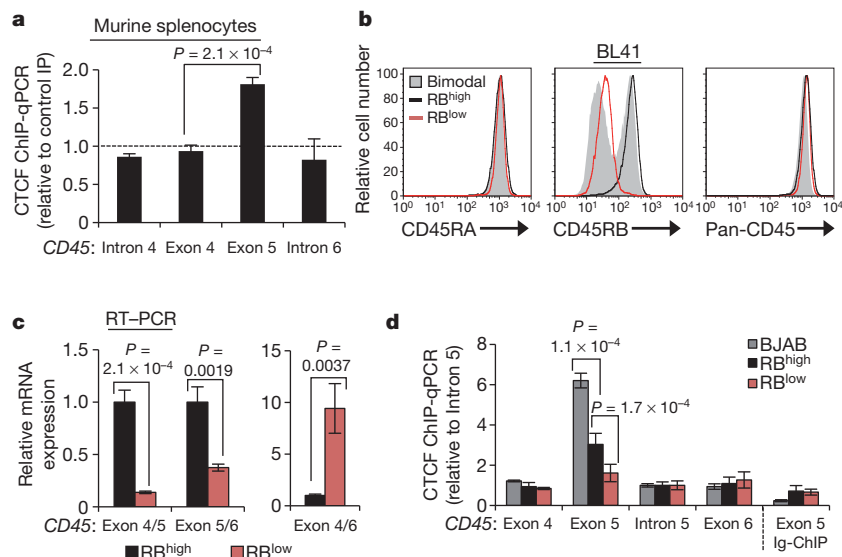


Figure 1 | Binding of CTCF to exon 5 of *CD45* DNA is associated with inclusion of exon 5 in *CD45* transcripts. **a**, CTCF ChIP in murine splenocytes and quantitative PCR (qPCR) relative to rabbit Ig control ChIP ($n = 2$). **b**, Cell-surface staining for CD45RA (exon 4-containing) and CD45RB (exon 5-containing) isoforms and total CD45 (pan) in parental BL41 B cells (RB^{high}), cell-culture-derived CD45RB bimodal cells, and CD45RB low (RB^{low}) cells

Whereas lymphocyte cell lines generally express high levels of CD45RB (Supplementary Fig. 2d, e), culturing BL41 cells in non-heat-inactivated fetal bovine serum (FBS) resulted in bimodal CD45RB expression. CD45RB low cells (RB^{low}) were sorted from the bimodal population and stably maintained (Fig. 1b). Parental BL41 cells (RB^{high}) and sorted RB^{low} cells express equivalent exon 4-containing CD45RA isoforms and total CD45 (Fig. 1b), indicating specific exclusion of exon 5 in the RB^{low} population. Quantitative RT-PCR with exon junction spanning primers validated exon 5 skipping: RB^{low} cells showed reduced exon 4/5 and exon 5/6 junctions, but enhanced exon 4/6 junctions relative to RB^{high} cells (Fig. 1c). Notably, several histone modifications that have been previously linked to alternative splicing (H3K36me3, H3K27me3, H3K4me3)^{16,29} are equivalently detected at exon 5 in RB^{high} and RB^{low} cells (Supplementary Fig. 3a, b). CTCF-ChIP in the newly identified RB^{high} and RB^{low} BL41 cells, and CD45RB-high BJAB cells revealed a strong positive correlation between exon 5 inclusion in *CD45* mRNA and CTCF binding at *CD45* exon 5 DNA (Fig. 1b–d and Supplementary Fig. 2e), particularly in BJAB cells, which also express elevated CTCF protein (Supplementary Fig. 3c). In agreement with the observation that exon 5 splicing is independent of hnRNPLL, modulation of hnRNPLL expression did not influence CTCF binding to *CD45* exon 5 (Supplementary Fig. 3d).

To assess whether the association between CTCF binding and exon 5 inclusion reflects a direct role for CTCF in *CD45* alternative splicing, we used RNA interference to deplete CTCF from our B cell lines (Supplementary Fig. 4a). Decreasing CTCF levels in bimodal BL41 cells led to a marked loss of CD45RB expression without reducing overall CD45 levels (Fig. 2a and Supplementary Fig. 4b). Similarly, CTCF depletion in RB^{low} cells and BJAB cells led to a substantial loss of CD45RB staining with little effect on overall CD45 levels (Fig. 2a). Quantitative RT-PCR of *CD45* mRNA in CTCF-depleted RB^{low} and BJAB cells validated reduced exon 5 expression and increased exon 4/6 junctions (Fig. 2b, c, respectively; Supplementary Fig. 4c, additional transductions), confirming that CTCF mediates exon 5 inclusion.

CTCF promotes pol II pausing at *CD45* exon 5

We next investigated the mechanism by which CTCF binding to *CD45* DNA influences mRNA splicing outcomes. Given that genome-wide ChIP-seq studies have revealed overlapping intragenic CTCF and pol

II peaks³⁰, we examined whether CTCF promotes inclusion of exon 5 through interference with pol II elongation. ChIP confirmed significant enrichment of pol II at *CD45* exon 5 DNA, but not at adjacent regions in RB^{high} cells as compared to RB^{low} cells (Fig. 3a). Using antibodies specific to pol II phosphorylated on the carboxy-terminal domain (CTD), we further showed that elevated pol II at *CD45* exon 5 in RB^{high} cells is associated with the elongating form phosphorylated on serine 2 of CTD YSPTSPS heptad repeats³¹ (Supplementary Fig. 5a, b). Notably, CTCF depletion from RB^{high} cells (Supplementary Fig. 5c, d) reduced both CTCF binding (Fig. 3b) and pol II levels at *CD45* exon 5 (Fig. 3c).

II peaks³⁰, we examined whether CTCF promotes inclusion of exon 5 through interference with pol II elongation. ChIP confirmed significant enrichment of pol II at *CD45* exon 5 DNA, but not at adjacent regions in RB^{high} cells as compared to RB^{low} cells (Fig. 3a). Using antibodies specific to pol II phosphorylated on the carboxy-terminal domain (CTD), we further showed that elevated pol II at *CD45* exon 5 in RB^{high} cells is associated with the elongating form phosphorylated on serine 2 of CTD YSPTSPS heptad repeats³¹ (Supplementary Fig. 5a, b). Notably, CTCF depletion from RB^{high} cells (Supplementary Fig. 5c, d) reduced both CTCF binding (Fig. 3b) and pol II levels at *CD45* exon 5 (Fig. 3c).

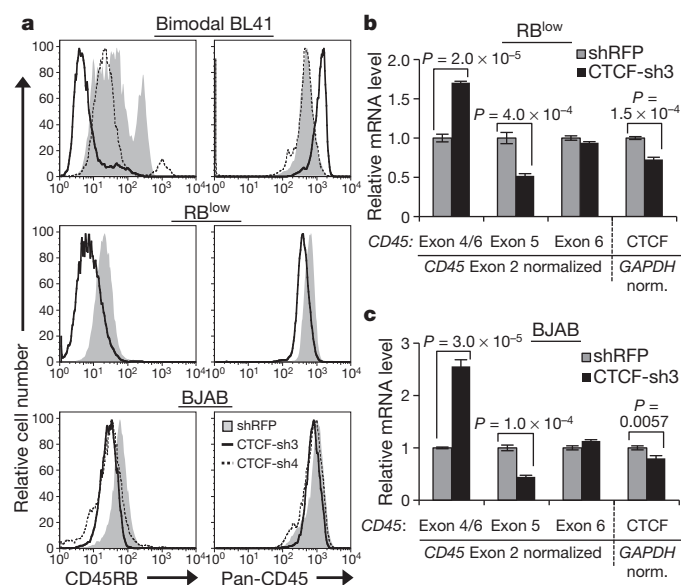


Figure 2 | CTCF depletion leads to reduced exon 5 inclusion in *CD45* transcripts. **a**, Cell-surface CD45RB isoform and total CD45 expression in cells transduced with short hairpin RNA (shRNA) against CTCF (CTCF-sh3 and/or sh-4) or control shRNA against red fluorescent protein (RFP). **b**, **c**, qRT-PCR in CTCF-depleted RB^{low} (**b**) and BJAB cells (**c**) from **a** to detect *CD45* (left) and *CTCF* (right) mRNA levels ($n = 3$). Graphs show mean values \pm s.d. P , two-tailed Student's t test.

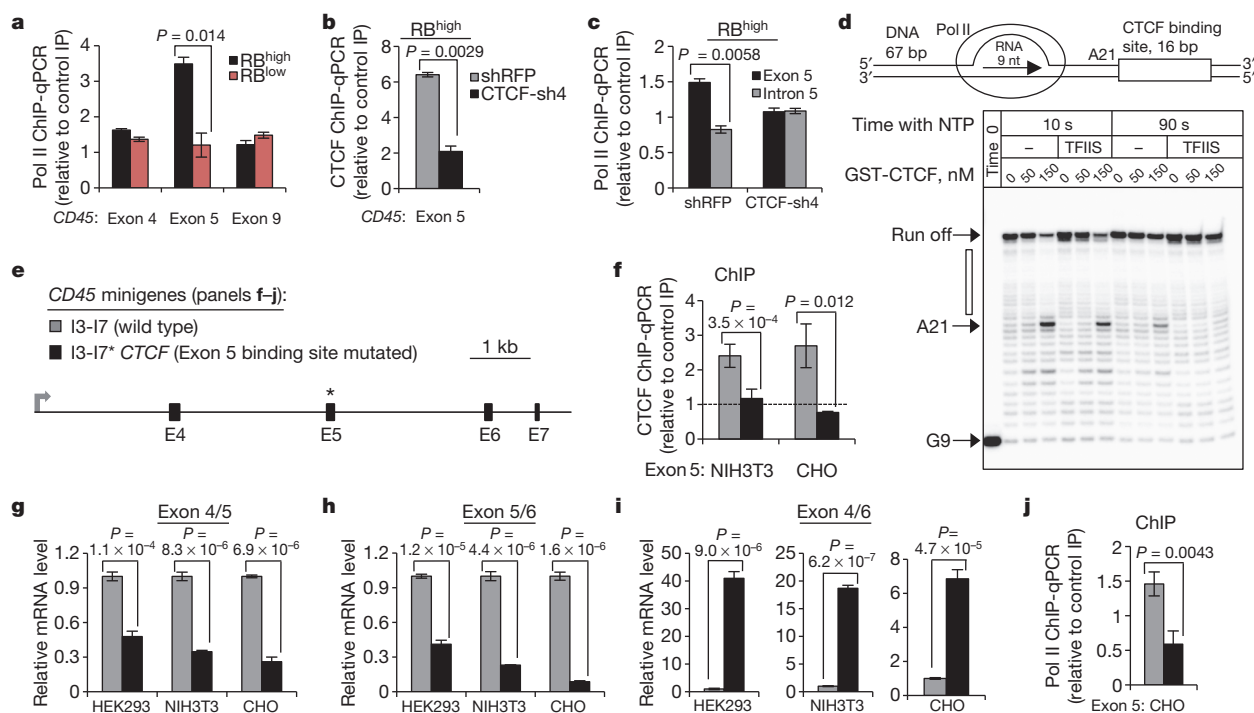


Figure 3 | CTCF binding at *CD45* exon 5 DNA facilitates exon 5 inclusion in *CD45* transcripts through local pol II pausing. **a**, RNA pol II ChIP and qPCR relative to mouse Ig control IP ($n = 3$). **b**, CTCF ChIP in RB^{high} cells transduced with shRNA against *CTCF* versus shRFP-transduced cells and qPCR relative to rabbit Ig control IP ($n = 2$). **c**, RNA pol II ChIP of RB^{high} cells from **b** and qPCR relative to mouse Ig control IP ($n = 2$). **d**, *In vitro* transcription with a DNA oligo incorporating a CTCF binding site at position 26 relative to elongation complex assembly. Recombinant CTCF and TFIIS protein were introduced as indicated, with variable effects on pausing at adenine 21 (A21).

The above data definitely link CTCF binding, pol II pausing and exon 5 inclusion, but do not exclude additional, context-dependent secondary effects. To query whether CTCF binding to an actively transcribed template is sufficient to promote pol II pausing, we assembled a pol II ternary elongation complex from synthetic DNA and RNA oligonucleotides and highly purified yeast pol II²². A CTCF binding site was incorporated into the template DNA at position 26 relative to the hybridization location of a 9-nucleotide RNA primer (Fig. 3d). CTCF binding to the target sequence was confirmed by electrophoretic mobility shift assay (EMSA) (Supplementary Fig. 6a). Incubation with pol II and increasing amounts of recombinant CTCF resulted in pausing immediately upstream of the CTCF binding site (Fig. 3d). Extended incubation or introduction of the elongation factor TFIIS substantially reduced pausing and led to near complete escape of paused pol II (Fig. 3d and Supplementary Fig. 6b). Thus, CTCF can autonomously promote pol II pausing, but not complete arrest, on a naked DNA template. These data establish CTCF as a direct impediment to transcription that can act in the absence of a particular nucleosome structure or chromatin context. Furthermore, the ability of paused pol II to resume transcription efficiently in the presence of CTCF supports a physiological role for CTCF in favouring exon inclusion through transient, spatiotemporal pol II pausing.

Having demonstrated that CTCF can promote pol II pausing, we explored the relationship between CTCF, pol II and exon inclusion in a tractable, endogenous system. We generated a wild-type minigene extending from intron 3 through intron 7 of human *CD45* genomic DNA (I3-I7), as well as a mutant analogue, in which the exon 5 CTCF binding site was disrupted through nucleotide substitution (I3-I7**CTCF*) (Fig. 3e and Supplementary Fig. 6c). The 11 zinc fingers of CTCF support multiple contacts to substrate DNA^{33,34} and a minimum of five substitutions within the core motif²⁵ were required to

e, Representation of *CD45* minigenes with wild-type (I3-I7) or mutated exon 5 CTCF binding site (I3-I7**CTCF*), used in **f–j**. **f**, CTCF-ChIP in NIH3T3 and CHO cells transfected with the *CD45* minigenes and qPCR relative to rabbit Ig control IP. Error bars represent standard error of the mean (s.e.m.) ($n = 3$). **g–i**, qRT-PCR from minigene-transfected HEK293, NIH3T3 and CHO cells to detect the junctions of exons 4/5 (**g**), 5/6 (**h**) and 4/6 (**i**) relative to exon 6 ($n = 3$). **j**, RNA pol II ChIP in CHO cells transfected with the *CD45* minigenes and qPCR relative to mouse Ig control IP ($n = 3$). Unless indicated otherwise, graphs show mean values \pm s.d. P , two-tailed Student's t test.

significantly ablate CTCF binding (EMSA, Supplementary Fig. 6d). To avoid detection of endogenous *CD45*, which is confined to the haematopoietic lineage¹⁸, the minigenes were transfected into several fibroblast cell lines. In addition to human HEK293 cells, murine NIH3T3 and hamster CHO cells were used to specifically amplify human minigene *CD45* DNA in ChIP analyses. CTCF ChIP of transfected NIH3T3 and CHO cells confirmed robust binding to exon 5 of the I3-I7 minigene, and complete disruption of binding to the mutated, I3-I7**CTCF* minigene (Fig. 3f). Quantitative RT-PCR indicated that both minigenes were comparably expressed and approximated endogenous *CD45* levels in immune cells (Supplementary Fig. 6e). Mutation of the CTCF binding site in exon 5 led to a marked decrease in 4/5 and 5/6 junctions, and increase in 4/6 junctions in all three cell types (Fig. 3g, h, i, respectively), resulting in an overall 50–100 \times decrease in exon 5 inclusion. Notably, ChIP confirmed increased pol II occupancy at exon 5 in the I3-I7 minigene, but not in the mutated, I3-I7**CTCF* minigene (Fig. 3j). As the two minigenes are identical in every regard minus the five core nucleotides of the CTCF binding site, these data establish CTCF as a direct regulator of *CD45* exon 5 inclusion, which operates through promoting local pol II pausing.

DNA methylation inhibits exon 5 CTCF binding

Armed with the knowledge that CTCF binding to exon 5 DNA regulates inclusion, and given that exon 5 is variably excluded during lymphocyte maturation, we asked whether and how CTCF binding is modulated to influence splicing outcome. Whereas CTCF is ubiquitously expressed, binding to DNA is inhibited by methylation on CpG dinucleotides^{27,34}. Several recent studies have shown that DNA methylation is substantially enriched at exons relative to introns^{14,15,36}, suggesting a role in pre-mRNA processing, yet a causal relationship between these processes has not been demonstrated. Methylated

DNA immunoprecipitation (MedIP) in our B-cell lines suggested that CTCF binding at *CD45* exon 5 and associated exon inclusion are indeed regulated by DNA methylation: we detected a strong inverse correlation between CTCF and 5-methylcytosine at *CD45* exon 5, but not at adjacent exons (Figs 1d, 4a). To assess whether DNA methylation of *CD45* exon 5 and reciprocal loss of CTCF binding contribute to exon 5 exclusion during the transition from naïve to mature T lymphocytes, $CD3^+$ T cells were isolated from human peripheral blood and sorted into RB^{high} (naïve) and RB^{medium} (mature) populations¹⁹ (Fig. 4b). MedIP confirmed significant enrichment of *CD45* exon 5 methylation (Fig. 4c) and reduced CTCF binding (Fig. 4d) in RB^{medium} cells compared to RB^{high} peripheral T cells. Thus, CTCF binding and *CD45* exon 5 inclusion are inversely related to DNA methylation in several transformed cell lines and in primary T cells.

To determine whether dynamic methylation of *CD45* exon 5 DNA is a regulatory mechanism contributing to *CD45* alternative splicing, we modulated methylation through inhibition of the DNA maintenance methyltransferase, DNMT1. We reasoned that, if elevated exon 5 methylation and consequent reduced CTCF binding were the principal components distinguishing RB^{low} and RB^{high} cells, inhibition of methylation should cause RB^{low} cells to revert to an RB^{high} phenotype. Indeed, DNMT1 depletion in RB^{low} cells (Supplementary Fig. 7a, b) reduced 5-methylcytosine levels (Fig. 4e) and restored CTCF binding at *CD45* exon 5 (Fig. 4f), leading to enhanced exon 5 inclusion in *CD45* mRNA, as evidenced by increased exon 4/5 and 5/6 junctions (Supplementary Fig. 7c) and cell-surface CD45RB (Fig. 4g). Notably, increasing CTCF binding in RB^{low} cells through reduced exon 5 methylation also reinstated local pol II pausing (Fig. 4h). In addition to identifying dynamic DNA methylation as a possible regulatory mechanism governing *CD45* alternative splicing *in vivo*, these data establish CTCF as the first mechanistic link between DNA methylation and alternative pre-mRNA splicing.

Global effects of intragenic CTCF on splicing

Although studies of CTCF function have been largely restricted to intergenic activities, CTCF ChIP-seq studies found that approximately 40–45% of CTCF binding sites are located intragenically^{35,37,38}. Based on our observations with *CD45*, we propose that some portion of intragenic CTCF binding sites operate to influence pre-mRNA processing decisions. To globally address the impact of CTCF on alternative splicing, we performed CTCF ChIP-seq in BL41 and

BJAB cells to produce cell-type-specific CTCF binding maps, and high-throughput RNA-sequencing (RNA-seq) of total RNA from CTCF-depleted BL41 and BJAB cells and their relevant controls (CTCF-sh3, Fig. 2a). Mapping of overall CTCF binding sites in BL41 and BJAB cells indicated comparable distribution patterns to previous reports (Supplementary Table 2). The mixture of isoforms (MISO) model was applied to RNA-seq data (Supplementary Table 3) to identify exons with a high probability of differential expression in response to CTCF depletion, as assessed by the Bayes factor confidence index³⁹. Exons showing altered inclusion in response to CTCF depletion were further subdivided into three categories based on proximity to a local CTCF binding site: unbound by CTCF, or CTCF-bound within 1 kilobase downstream or upstream of the exon (Fig. 5a and Supplementary Table 4). CTCF is a global regulator of transcription²⁷ and depletion would be expected to result in some level of alternative splicing due to alterations in upstream pathways. Accordingly, MISO identified exons that were differentially included in mRNA in response to CTCF depletion, but were not locally bound by CTCF on the corresponding DNA. Importantly, in BL41 and BJAB cells, alternative exons not bound by CTCF were centred at zero across Bayes factor thresholds, indicating that secondary effects of CTCF depletion showed no preference towards exon inclusion or exclusion (Fig. 5b, c, respectively). Similarly, CTCF binding upstream of the differentially expressed exon did not show a statistically significant bias towards exon inclusion or exclusion (Fig. 5b, c and Supplementary Fig. 8a, b). However, we detected a strong correlation between CTCF depletion and exon exclusion if CTCF is bound downstream of the alternative exon in both BL41 and BJAB cells (Fig. 5b, c and Supplementary Fig. 8a, b). We additionally identified CTCF-bound exons that showed reduced inclusion in BL41 and BJAB cells, as well as unique examples, indicating a degree of cell-type specificity (Supplementary Figs 8c, 9a and Supplementary Table 4).

These genome-level data are consistent with our observations in the *CD45* model system, wherein CTCF binding downstream of the weak 3' splice site flanking exon 5 promoted inclusion of exon 5 in mature message, but had no effect on exon 6 (Figs 1c, d and 2b, c). As we had mechanistically linked CTCF-associated pol II pausing to *CD45* exon 5 inclusion, we examined pol II occupancy at the downstream CTCF sites that led to reduced exon inclusion upon CTCF depletion. Inspection of publicly available CTCF ChIP-seq data from $CD4^+$ T cells³⁷ indicated high conservation of these CTCF binding sites²⁶.

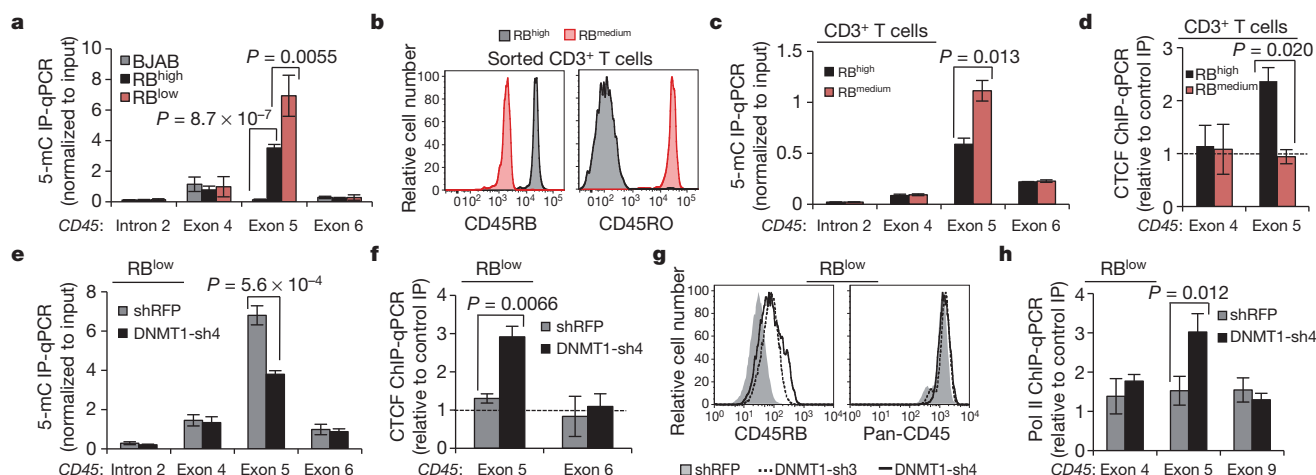


Figure 4 | 5-methylcytosine levels (5-mC) are inversely related to CTCF binding and exon 5 inclusion. **a**, Methylated DNA immunoprecipitation (MedIP) in B cell line genomic DNA and qPCR relative to input ($n = 5$). **b**, Representative *CD45* isoform expression in primary peripheral human $CD3^+$ T cells sorted on the basis of cell-surface CD45RB and CD45RO. **c**, MedIP and qPCR relative to input in sorted primary human $CD3^+$ T cells ($n = 6$, compiled from two donors). **d**, CTCF-ChIP and qPCR relative to rabbit

Ig control IP, in sorted primary $CD3^+$ T cells ($n = 2$). **e**, MedIP and qPCR relative to input in BL41 RB^{low} cells transduced with shRNA against *DNMT1* versus shRFP-transduced cells ($n = 3$). **f**, CTCF ChIP in cells from **e** and qPCR relative to rabbit Ig control IP ($n = 3$). **g**, Cell-surface CD45RB expression in cells from **e**. **h**, RNA pol II ChIP and qPCR in cells from **e** relative to mouse Ig control IP ($n = 3$). Unless indicated otherwise, graphs show mean values \pm s.d. *P*, two-tailed Student's *t* test.

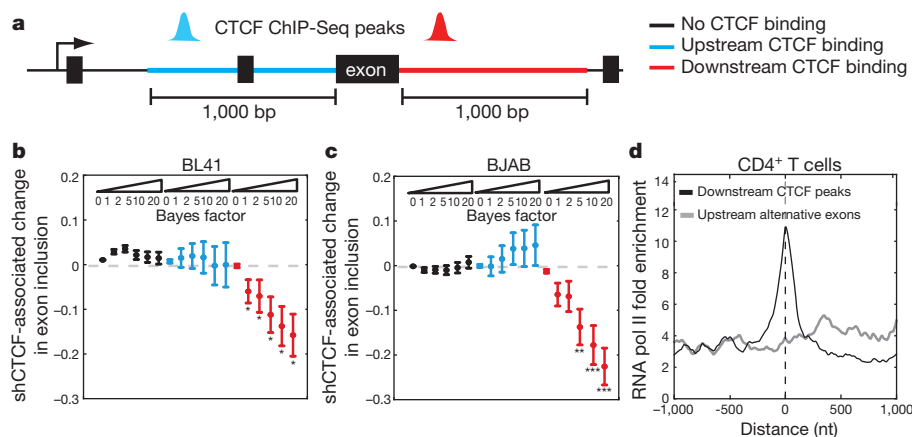


Figure 5 | Global identification of CTCF-dependent exons. **a**, Alternative exons were classified on the basis of the relative location of an exclusive CTCF peak within 1 kb of the exon. **b**, Difference in the mean exon inclusion level between bimodal BL41 cells transduced with shRNA against CTCF versus shRFP-transduced cells (from Fig. 2a) for exons with CTCF peak in upstream (blue) or downstream regions (red) but not in the exon body and for exons with no CTCF binding (black). The mean \pm s.e.m. for each class of exons is plotted

Analysis of the corresponding pol II ChIP-seq data³⁷ revealed a stronger enrichment of pol II at downstream CTCF binding sites relative to upstream exons (Fig. 5d). Enrichment of pol II occupancy at CTCF binding sites compared to associated upstream alternative exons was confirmed for several genes in BJAB and BL41 cells (Supplementary Fig. 9b). Together with our CD45 data, we conclude that CTCF bound downstream of alternative exons promotes pol II pausing, providing the necessary temporal context for co-transcriptional spliceosome assembly at weak upstream splice sites.

Discussion

In recent years, the link between DNA structure and pre-mRNA processing has been gaining increasing attention. Reports of increased nucleosome occupancy and DNA methylation as well as distinct histone methylation patterns at exons relative to introns have fuelled the hypothesis that exons are differentially marked to aid the spliceosome in the process of exon definition^{16,17}. It has further been shown that pol II occupancy increases in the vicinity of exons, although whether a function of DNA sequence, chromatin structure or the presence of DNA-binding proteins has not been defined. Recently, several studies have linked modification of distinct histone methylation patterns to alternative splicing^{16,17}. However, exonic histone methylation was shown to be equivalent in other models of robust exon inclusion versus exclusion⁴⁰, suggesting that, although histone methylation patterns may prime splicing decisions, they probably do so in concert with other factors. Consistent with the latter, we observed comparable histone methylation at exon 5 whether or not exon 5 was included in the CD45 message (Supplementary Fig. 3a, b). Rather, we show that mutually exclusive DNA methylation and CTCF binding regulate exon 5 inclusion through influencing pol II elongation dynamics (Supplementary Fig. 10). Given that mapping of CTCF binding sites shows roughly 40–70% conservation between tissues²⁷, it is tempting to speculate that altered DNA methylation patterns during development can lead to variations in intragenic CTCF binding that thereby contribute to tissue-specific alternative splicing patterns. This may be especially relevant in pathological conditions, such as cancer, where widespread changes in DNA methylation, altered CTCF binding, and aberrant alternative pre-mRNA splicing have been reported^{15,41–43}. We predict that our identification of CTCF as a DNA-binding regulator of alternative pre-mRNA splicing represents the tip of the iceberg, and that a long list of location-specific DNA-binding ‘splicing factors’ will follow.

against increasing Bayes factor thresholds. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, Wilcoxon rank sum test for differences in exon inclusion at the different thresholds. **c**, Same as **b** for BJAB shCTCF compared with wild-type BJAB cells (from Fig. 2a). **d**, Normalized CD4⁺ T cell RNA pol II read signal centred on the alternative exon or the corresponding downstream CTCF peak summit.

METHODS SUMMARY

Experiments were performed with BJAB and BL41 cells or primary lymphocytes. CD45 isoform analysis was achieved with isoform-specific antibodies or pan-antibody directed against a common region of CD45. Transductions were executed with vesicular stomatitis virus G (VSV-G)-pseudotyped lentivirus, and selected for puromycin resistance. Quantitative RT-PCR was performed on cDNA from total RNA. Protein lysates were prepared with RIPA buffer. ChIP and MedIP were conducted with formaldehyde cross-linked, sonicated material. *In vitro* transcription elongation was performed with yeast RNA pol II, yeast TFIIS and human CTCF. Minigenes were cloned into the pCI-neo (Promega) construct and transfected with Lipofectamine 2000 (Invitrogen). ChIP-Seq and RNA-Seq were executed with the Illumina platform. For ChIP-Seq, Illumina FastQ files were mapped to the human genome (hg19). Peak calling was run using Rabbit Ig control sequencing data as background. For RNA-Seq, exon inclusion levels were determined using the MISO program³⁹.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 March; accepted 12 August 2011.

Published online 2 October; corrected 3 November 2011 (see full-text HTML version for details).

- Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genet.* **40**, 1413–1415 (2008).
- Tazi, J., Bakkour, N. & Stamm, S. Alternative splicing and disease. *Biochim. Biophys. Acta* **1792**, 14–26 (2009).
- Matlin, A. J., Clark, F. & Smith, C. W. Understanding alternative splicing: towards a cellular code. *Nature Rev. Mol. Cell Biol.* **6**, 386–398 (2005).
- Han, J., Xiong, J., Wang, D. & Fu, X. D. Pre-mRNA splicing: where and when in the nucleus. *Trends Cell Biol.* **21**, 336–343 (2011).
- Singh, R. & Valcarcel, J. Building specificity with nonspecific RNA-binding proteins. *Nature Struct. Mol. Biol.* **12**, 645–653 (2005).
- Kornblihtt, A. R. Coupling transcription and alternative splicing. *Adv. Exp. Med. Biol.* **623**, 175–189 (2007).
- de la Mata, M. *et al.* A slow RNA polymerase II affects alternative splicing *in vivo*. *Mol. Cell* **12**, 525–532 (2003).
- Tilgner, H. *et al.* Nucleosome positioning as a determinant of exon recognition. *Nature Struct. Mol. Biol.* **16**, 996–1001 (2009).
- Spies, N., Nielsen, C. B., Padgett, R. A. & Burge, C. B. Biased chromatin signatures around polyadenylation sites and exons. *Mol. Cell* **36**, 245–254 (2009).
- Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C. & Komorowski, J. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res.* **19**, 1732–1741 (2009).
- Kolasinska-Zwier, P. *et al.* Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature Genet.* **41**, 376–381 (2009).
- Schwartz, S., Meshorer, E. & Ast, G. Chromatin organization marks exon-intron structure. *Nature Struct. Mol. Biol.* **16**, 990–995 (2009).

14. Chodavarapu, R. K. *et al.* Relationship between nucleosome positioning and DNA methylation. *Nature* **466**, 388–392 (2010).
15. Hodges, E. *et al.* High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. *Genome Res.* **19**, 1593–1605 (2009).
16. Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R. & Misteli, T. Epigenetics in alternative pre-mRNA splicing. *Cell* **144**, 16–26 (2011).
17. All6, M. *et al.* Chromatin and alternative splicing. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 103–111 (2010).
18. Trowbridge, I. S. & Thomas, M. L. CD45: an emerging role as a protein tyrosine phosphatase required for lymphocyte activation and development. *Annu. Rev. Immunol.* **12**, 85–116 (1994).
19. Hermiston, M. L., Xu, Z. & Weiss, A. CD45: a critical regulator of signaling thresholds in immune cells. *Annu. Rev. Immunol.* **21**, 107–137 (2003).
20. Oberdoerffer, S. *et al.* Regulation of CD45 alternative splicing by heterogeneous ribonucleoprotein, hnRNPL. *Science* **321**, 686–691 (2008).
21. Topp, J. D., Jackson, J., Melton, A. A. & Lynch, K. W. A cell-based screen for splicing regulators identifies hnRNP L as a distinct signal-induced repressor of CD45 variable exon 4. *RNA* **14**, 2038–2049 (2008).
22. Motta-Mena, L. B., Heyd, F. & Lynch, K. W. Context-dependent regulatory mechanism of the splicing factor hnRNP L. *Mol. Cell* **37**, 223–234 (2010).
23. Tong, A., Nguyen, J. & Lynch, K. W. Differential expression of CD45 isoforms is controlled by the combined activity of basal and inducible splicing-regulatory elements in each of the variable exons. *J. Biol. Chem.* **280**, 38297–38304 (2005).
24. Horgan, K. J. *et al.* CD45RB expression defines two interconvertible subsets of human CD4⁺ T cells with memory function. *Eur. J. Immunol.* **24**, 1240–1243 (1994).
25. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
26. UCSC Genome Browser, GRC37/hg19, ENCODE Histone Modification Tracks.
27. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194–1211 (2009).
28. Ohlsson, R., Bartkuhn, M. & Renkawitz, R. CTCF shapes chromatin by multiple mechanisms: the impact of 20 years of CTCF research on understanding the workings of chromatin. *Chromosoma* **119**, 351–360 (2010).
29. Luco, R. F. *et al.* Regulation of alternative splicing by histone modifications. *Science* **327**, 996–1000 (2010).
30. Wada, Y. *et al.* A wave of nascent transcription on activated human genes. *Proc. Natl Acad. Sci. USA* **106**, 18357–18361 (2009).
31. Buratowski, S. Progression through the RNA polymerase II CTD cycle. *Mol. Cell* **36**, 541–546 (2009).
32. Komissarova, N., Kireeva, M. L., Becker, J., Sidorenkov, I. & Kashlev, M. Engineering of elongation complexes of bacterial and yeast RNA polymerases. *Methods Enzymol.* **371**, 233–251 (2003).
33. Filippova, G. N. *et al.* An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol. Cell. Biol.* **16**, 2802–2813 (1996).
34. Renda, M. *et al.* Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. *J. Biol. Chem.* **282**, 33336–33345 (2007).
35. Kim, T. H. *et al.* Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**, 1231–1245 (2007).
36. Lyko, F. *et al.* The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol.* **8**, e1000506 (2010).
37. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
38. Jothi, R., Cuddapah, S., Barski, A., Cui, K. & Zhao, K. Genome-wide identification of *in vivo* protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res.* **36**, 5221–5231 (2008).
39. Katz, Y., Wang, E. T., Airolidi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods* **7**, 1009–1015 (2010).
40. Huff, J. T., Plocik, A. M., Guthrie, C. & Yamamoto, K. R. Reciprocal intronic and exonic histone modification regions in humans. *Nature Struct. Mol. Biol.* **17**, 1495–1499 (2010).
41. Berdasco, M. & Esteller, M. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. *Dev. Cell* **19**, 698–711 (2010).
42. David, C. J. & Manley, J. L. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev.* **24**, 2343–2364 (2010).
43. Jelinic, P. & Shaw, P. Loss of imprinting and cancer. *J. Pathol.* **211**, 261–268 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Rao, C. Burge and K. Lynch for critical reading of this manuscript. We also thank A. Rao for reagents and K. Nyswaner and M. Prigge for technical assistance. This work is supported by the Intramural Research Program of NIH, the National Cancer Institute, The Center for Cancer Research (S.O., P.O., M.K.), and the Swedish Research Council Foundation and the Foundation for Strategic Research (R.S.).

Author Contributions S.S. performed ChIP, MedIP and EMSA. M.G. and S.S. performed lentiviral transductions, transfections, flow cytometry, and qPCR. E.K. analysed ChIP and RNA-seq data. M.I. performed *in vitro* transcription. S.S. and B.S. cloned the minigenes. All authors designed experiments and M.K., P.O., R.S. and S.O. supervised the project. S.O. and R.S. wrote the text. P.O. edited the text.

Author Information All data sets in this publication are available in the NCBI Gene Expression Omnibus accession number GSE31278. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.O. (shalini.oberdoerffer@nih.gov).

METHODS

Cell culture. BJAB and BL41 cells were maintained at 37 °C, 5% CO₂ in RPMI (Invitrogen) supplemented with 10% FBS (Hyclone), and 1% L-glutamine. BJAB and parental RB^{high} BL41 cells were cultured in heat-inactivated FBS, whereas RB^{low} BL41 cells were initially kept in native FBS, but were ultimately transitioned into inactivated serum. JSL1 cells were maintained at 37 °C, 5% CO₂ in RPMI (Invitrogen) supplemented with 5% FBS (Hyclone), and 1% L-glutamine. Primary human peripheral blood lymphocytes were purified by spinning through Ficoll Paque (GE Healthcare). Isolated cells were washed twice with PBS and CD3⁺ T cells were isolated with CD3⁺ microbeads (Miltenyi Biotec). Primary murine splenocytes were isolated from whole spleen of BL/6 mice. Single-cell suspensions were lysed with ACK lysis buffer (0.15 M NH₄Cl, 10 mM KHCO₃, 0.1 mM Na₂EDTA) to remove red blood cells before ChIP assay. JSL1 cells were stimulated at a concentration of 3 × 10⁵ cells per ml. Phorbol 12-myristate 13-acetate (PMA) was added at a final concentration of 20 nM. Flow cytometry was performed 2 days post-stimulation.

Virus production. Constructs encoding shRNA directed against *CTCF* and *DNMT1* were obtained from Open Biosystems and were transfected (Lipofectamine 2000, Invitrogen) along with VSV-G and *gag/pol* (courtesy of The RNAi Consortium of the Broad Institute) into 293T cells for viral production. Viral supernatants were concentrated 50× and aliquoted for storage.

Cell line infection. BJAB and BL41 cells were plated in 96-well round-bottom plates at 100,000 cells per well. Five microlitre of virus and 8 µg ml⁻¹ polybrene were added per well and the plate was spun at 760g for 90 min. The supernatants were removed and fresh media was added. Puromycin was added at a final concentration of 5 µg ml⁻¹ on day 2. Depletion of CTCF from cells resulted in significant cell death after 1 week in culture and depletion of DNMT1 resulted in silencing after 10 days in culture. Cells for downstream analysis were collected 5 (shCTCF) or 7 (shDNMT1) days post-infection. To scale up infections for ChIP and western blotting, infections were performed in individual wells of 96-well plates and pooled before harvesting for RNA, ChIP and western blot. Three plates were pooled for shCTCF experiments and 3.5 plates were pooled for shDNMT1 experiments. Three individual RNA and ChIP samples were taken from each of the bulk cultures.

Target sequences of shRNAs. DNMT1-sh3, 5'-CGAGAAGAATATCGAAC TCTT-3'; DNMT1-sh4, 5'-CGACTACATCAAAGGCAGCAA-3'; CTCF-sh3, 5'-CCTCCTGAGGAATCACCTTAA-3'; CTCF-sh4, 5'-GCGGAAAGTGAA CCCATGATA-3'; shRFP, 5'-GAATTAAGAGAGGCTCAGTTA-3'; LL-sh4, 5'-CGACAGGCTCTAGTGAATTT-3'.

Flow cytometry. The following antibodies were used for flow cytometry: CD45RO clone UCHL1 (eBioscience, 12-0457-42, batch no. E034572), CD45RA clone MEM-56 (ExBio, 1P-223, batch no. 11827), CD45RB clone MT4 (BD Pharmingen, 555904, batch no. 89956) and pan-CD45 clone HI30 (BD Pharmingen, 555483, batch no. 555483). Staining of CD45 isoforms was performed in separate tubes, to avoid competition for antibody binding. Flow cytometry was performed on either a BD FACSCalibur or BD LSR II cytometer.

Quantitative RT-PCR. RNA was isolated with the Qiagen RNeasy Mini Kit and reverse transcription was performed with SuperScript II (Invitrogen) according to the manufacturer's instructions. PCR measurements were performed in triplicate in the presence of SYBR green reagent (Roche) and amplification was performed on a 480 Light Cycler (Roche). The average cycle thresholds for the technical triplicates were calculated to yield one value per primer set for each biological replicate. Normalization was performed to *GAPDH*, *RPS16* or surrounding exon level values using the formula $2^{(C_{\text{normalization}} - C_{\text{experimental}})}$ to determine relative expression. Averages and standard deviations of the normalized biological replicate values were plotted in the figures and used in *t*-test calculations. Figure legends indicate the number of biological replicates (individual RNA preparations) used in each experiment.

Western blots. Cells were lysed in RIPA buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1% NP40, 0.1% SDS, 0.5% sodium deoxycholate, and 1× Halt protease inhibitor cocktail (Thermo Scientific)). Proteins (35 µg) were loaded per lane on a 4–20% gradient SDS-PAGE gel. Western blot was performed with anti-CTCF clone D31H2 (Cell signaling 3418S, batch no. 1), DNMT1 antibody (Abcam ab13537, batch no. GR16960-1), or anti-p65 RelA (BD Bioscience 610869, batch no. 50886) antibodies. Anti-RelA immunoblotting served as a loading control for protein levels.

Chromatin immunoprecipitation (ChIP). Ten million cells were cross-linked for 10 min in 1% formaldehyde (Sigma) at room temperature, and quenched by adding glycine to a final concentration of 0.125 M for 5 min at room temperature. Cells were washed twice in chilled PBS, resuspended in buffer containing 50 mM HEPES-KOH (pH 7.5), 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100 and protease inhibitors (Thermo Scientific) and kept on ice for 10 min. Nuclei were pelleted at 800g for 5 min at 4 °C and resuspended in

buffer containing 10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and protease inhibitors (Thermo Scientific) followed by a 10-min incubation on ice. Nuclei were collected and resuspended in sonication buffer containing 10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5% EGTA, 0.1% Na-deoxycholate, 0.5% *N*-lauryl sarcosine and protease inhibitors (Thermo Scientific). Sonication of DNA was performed in an ultra sonicator water bath (Bioruptor) using two ten cycle runs of 30 s 'on' and 30 s 'off' to achieve an average fragment length of 200–400 bp. After addition of 1% Triton X-100, samples were centrifuged at 16000g for 10 min at 4 °C. An aliquot of sonicated DNA was reverse-crosslinked and run on a 1% agarose gel to confirm fragment size during each ChIP procedure. Chromatin (25 µg) was immunoprecipitated by adding the antibody of interest followed by overnight incubation at 4 °C. The following antibodies were used for ChIP: anti-CTCF (Millipore 07-729, batch no. DAM1772428), anti-RNA polymerase II clone 4H8 (Millipore 05-623, batch no. DAM1731474), anti-Ser2P RNA polymerase II clone H5 (Covance MMS129R, batch no. E10017AF), anti-Ser5P RNA polymerase II clone H14 (Covance MMS134R, batch no. E10142DF), anti-H3K36Me3 (Abcam ab9050, batch no.947467), anti-H3K27Me3 (Abcam ab6002, batch no. 934602), anti-H3K4Me3 clone MC315 (Millipore 04-745, batch no. NG1717145), Normal Rabbit IgG (Cell signaling Technology 2729, batch no. 4), and normal mouse IgG (Millipore 12-371, batch no. 1718089). After overnight incubation, 30 µl of Dynal Protein A/G beads (Invitrogen) or Protein L magnetic beads (Biovision) (for phosphorylated RNA Pol II antibodies) were added and incubated for 1 h at 4 °C. Beads were washed sequentially for 3 min each in low salt (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), high salt (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), LiCl buffer (10 mM Tris-HCl pH 8.0, 0.25 M LiCl, 1% NP40, 1% Na-deoxycholate) and TE buffer. Beads were eluted in 150 µl elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS, 50 mM NaHCO₃) and treated with 1 µl RNase A (1 mg ml⁻¹ Ambion) at 37 °C for 30 min. Cross-linking was reversed and proteins were degraded by addition of 1 µl proteinase K (20 mg ml⁻¹ Ambion) and incubation at 65 °C for 4 h. Eluted DNA was purified with QIAquick PCR purification (Qiagen), according to the manufacturer instructions.

Immunoprecipitated DNA and 5% input DNA were analysed by SYBR-Green real-time quantitative PCR. PCR measurements were performed in duplicate. The average cycle thresholds for the technical replicates were calculated to yield one value per primer set for each biological replicate and normalized to input using the formula $2^{(C_{\text{input}} - C_{\text{immunoprecipitation}})}$. These values were further normalized relative to the rabbit or mouse Ig control IP values for the primer set. Averages and standard deviations of the normalized biological replicate values were plotted in the figures and used in *t*-test calculations. Figure legends indicate the number of biological replicates (individual IPs) used in each experiment.

Methylated DNA immunoprecipitation (MedIP). MedIP was performed essentially according to the protocol described in ref. 44. Genomic DNA was purified from approximately 25 million cells using Zymo research Quick gDNA Midiprep kit (D3100), according to the manufacturer's instructions. For primary cells, CD3⁺ T cells were isolated from peripheral blood using CD3 microbeads (Miltenyi Biotec). CD3⁺ T cells were sorted into CD45RB high and CD45RB medium populations based on surface receptor staining of CD45RB and CD45RO. Purified genomic DNA was diluted into a total of 300 µl TE buffer and sonicated with a Bioruptor (10 cycles at low power, of 30 s 'on' and 30 s 'off') to an average size of 300–500 bp. An aliquot of sonicated DNA was run on 1% agarose gel to confirm fragment size during each MedIP procedure. Sonicated DNA (4 µg; 3 µg for primary cells) was denatured by incubation at 95 °C for 10 min and was immediately transferred to ice for 10 min. Immunoprecipitation buffer containing 10 mM sodium phosphate, 140 mM NaCl and 0.05% Triton X-100 was added to a final volume of 500 µl. For each IP reaction, 10 µg (8 µg for primary cells) of 5-methyl cytidine antibody clone b (Diagenode MAb-006-100, batch no. DA-0018) was added and incubated overnight at 4 °C with shaking. Five percent of DNA was kept as input.

After incubation, 30 µl of Dynal Protein G beads were added and further incubated for 1 h at 4 °C. Beads were washed thrice with 500 µl of IP buffer. Elution buffer (150 µl) containing 50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS, 50 mM NaHCO₃ and 20 µg proteinase K was added and incubated at 55 °C for 3 h. Tubes were applied to a magnetic rack and eluted DNA and input DNA were purified with the QIAquick PCR purification kit (Qiagen) followed by SYBR-Green real-time quantitative PCR to identify methylated regions. PCR measurements were performed in duplicate. The average cycle thresholds for the technical replicates were calculated to yield one value per primer set for each biological replicate and normalized to input using the formula $2^{(C_{\text{input}} - C_{\text{immunoprecipitation}})}$. Averages and standard deviations of the normalized biological replicate values were plotted in the figures and used in *t*-test calculations. Figure legends indicate the number of biological replicates (individual IPs) used in each experiment.

In vitro transcription elongation assay. RNA pol II from yeast *Saccharomyces cerevisiae* containing a histidine-tagged Rpb3 subunit was purified as described previously⁴⁵. Histidine-tagged TFIIS expression plasmid⁴⁶ was a gift from C. Kane. Recombinant TFIIS was purified according to ref. 46, with an additional purification on a Mono-S column (GE Healthcare). Human CTCF recombinant protein was obtained from Abnova (catalogue no. H00010664-P01, batch no. 0991020-2).

Elongation complex incorporating a 9-nt RNA was assembled as described previously⁴⁷, purified with Amicon Ultra-0.5 ml centrifugal filter (Millipore), and diluted with transcription buffer (TB; 20 mM Tris-HCl pH 7.9, 5 mM MgCl₂, 10 mM 2-mercaptoethanol, 40 mM KCl, 0.1 mg ml⁻¹ BSA). The reaction was initiated by mixing 5 µl of TEC +/- X µM CTCF with 5 µl of 0.1–0.5 mM NTP (GE Healthcare) +/- 1 µM TFIIS in TB and was terminated with gel-loading buffer (5 M urea, 25 mM EDTA at final concentration). RNA products were resolved in 20% denaturing polyacrylamide gels and visualized with a Typhoon 8600 phosphorimager (GE Healthcare).

Oligonucleotides used for elongation complex. Sequences of RNA and DNA oligonucleotides are as follows. RNA, 5'-AUCGAGAGG-3'; DNA with CTCF binding site, non-template strand, 5'-GGTATAGGATACTTACAGCCATCGA GAGGGACAAGGCGAAAGCATCCACCAGGGGCGCCAGCTAAT-3'; template strand, 5'-ATTAGCTGGCGCCCCCTGGTGATGCTTTCGCCTTGCTCCTCGATGGCTGTAAGTATCTATACCTAAT-3'.

Electrophoretic mobility shift assay (EMSA). The CTCF-binding oligonucleotides used for EMSA correspond to either the template used for *in vitro* transcription (Supplementary Fig. 6a) or the CTCF binding sites in the wild-type and mutated I3-I7 minigenes (Supplementary Fig. 6d). The two strands of DNA were annealed, 5' end-labelled with [γ -³²P] ATP and purified with a G-50 Micro column (GE healthcare). DNA probe (3 pM) equalling approximately 70,000 c.p.m. was mixed with glutathione S-transferase (GST)-tagged CTCF in binding buffer containing PBS and 5 mM MgCl₂, 0.1 mM ZnSO₄, 1 mM DTT, 0.1% NP40 and 10% glycerol. EMSA reaction mixtures (20 µl final volume) were incubated for 20 min at room temperature followed by electrophoresis on 5% native polyacrylamide gels and visualized as described above for *in vitro* transcription.

EMSA DNA probe sequences. *In vitro* transcription probe, 5'-CATCCACCAG GGGGCGCCAGCTAAT-3' and 5'-ATTAGCTGGCGCCCCCTGGTGATG-3'; wild-type exon 5 probe, 5'-TCAGTTCACAGAGGGCGTCTGCG-3' and 5'-CGCAGACGCCCTCTGCTGGAACCTGA-3'; mutated exon 5 probe, 5'-TCAG TTAAGCTGAGTACGTCTGCG-3' and 5'-CGCAGACGTAAGCTTAA CTGA-3'.

ChIP-Seq analyses. Illumina FastQ files were mapped to the human genome (hg19) using Bowtie⁴⁸ requiring a unique match (using the '-m 1' flag). The aligned reads in SAM format were converted to BED format before running the MACS peak caller⁴⁹. MACS peak calling was run using the rabbit Ig control sequencing data as background files for BJAB and BL41 CTCF ChIP-Seq data, respectively. The number of peaks identified per ChIP-Seq sample and sequenced reads are listed in Supplementary Table 2.

To investigate the effects of CTCF binding upon pre-mRNA splicing, we compared CTCF ChIP-Seq peaks with a set of alternative exons¹ requiring that the CTCF peak summit was located within 1,000 bp of the alternative exon boundaries (see Fig. 5a). Based on the presence of local CTCF peaks in BL41, BJAB and CD4 we classified each alternative exon into exons that were unbound by CTCF, and exons with either downstream or upstream CTCF binding. Classified unbound exons lacked CTCF peak summits in both the alternative exon body and within 1,000 bp on either side of the alternative exon. Exons with downstream CTCF binding had one or more CTCF peak summits within the region spanning from the alternative exon 5' splice site and 1,000 bp downstream in one or more of the CTCF data in BJAB, BL41 or CD4. Any alternative exons with a downstream CTCF peak but additional peak summits in the upstream region or within the alternative exon were not considered. The reciprocal procedure was used to classify exons with upstream CTCF binding. Alternative exons classified by local CTCF binding together with exon inclusion levels are provided in Supplementary Table 4.

RNA-Seq analyses. Illumina FastQ files were mapped to the human genome (hg19) and a collection of junctions using Tophat version 1.1.4 (ref. 50), using the paired-end mode and requiring a unique match. The resulting SAM file with uniquely mapped reads was converted to BAM format using samtools⁵¹. Mapping statistics for the RNA-Seq data are provided in Supplementary Table 3. We estimated exon inclusion levels of a collection of 42,557 alternative exons

(approximately the same as in ref. 1) using the MISO program³⁹ with the default parameters using the 'compute-genes-psi' function. The estimated exon inclusion levels from different RNA-Seq experiments were compared using MISO function 'compare-samples' to obtain exon inclusion level differences and Bayes factors. Statistically significant differences in exon inclusion levels between CTCF-bound and -unbound exons at different thresholds were evaluated using the Wilcoxon rank sum test. The overall difference in gene expression across RNA-Seq samples was evaluated using singular value decomposition. First we computed the expression level of each Refseq transcript as reads per kilobase and million mappable reads using the rpkmforgenes program⁵². The full gene expression matrix was normalized to unit length per transcript and subsequently used as input for singular value decomposition using the svdman program⁵³. The result in Supplementary Fig. 8c was obtained by projecting each sample onto the first two 'eigenarrays'⁵⁴.

RNA polymerase II ChIP-Seq analysis. We generated normalized RNA polymerase II fold enrichment signals over a set of regions by dividing the observed read sum at each position with the expected read sum computed as: (total_reads * read_length * number_of_regions) / genome_length. The normalized fold enrichment at each position was smoothed by window averaging using a window size of 100 nucleotides. We analysed all alternative exons with a downstream CTCF peak summit conserved in CD4⁺ T cells (Supplementary Table 4) after removing exons with a CTCF peak within 1,000 bp of an annotated transcript start site or poly A site in Ensembl (to remove effects from strong RNA pol II signals at transcript start and end locations). This procedure rendered 408 exons from which we computed both the CTCF peak summit position and exon middle coordinate. These two sets of 408 genomic coordinates each were used as the centre for the analysis in Fig. 5d. The same procedure was used to generate CTCF peak summits and middle exon positions for alternative exons with a conserved CTCF peak in the upstream region for Supplementary Fig. 8e (number of exons identified was 416).

Minigenes and transfection. CD45 minigenes were cloned into the pC1-neo mammalian expression vector (Promega). The wild-type CD45 minigene consists of 9.7 kb of CD45 genomic DNA sequence extending from 2.4 kb of intron 3 through 588 bp of intron 7. The I3-I7*CTCF minigene was made by mutating the CTCF binding site of exon 5 using site-directed mutagenesis with the primers indicated in Supplementary Table 1. The minigenes were transfected (Lipofectamine 2000, Invitrogen) along with pC1-neo vector control into HEK293 cells, CHO cells and NIH-3T3 cells. Cells were collected 48 h after transfection for RNA isolation (RNeasy, Qiagen) and chromatin immunoprecipitation. Transfection was performed in triplicate for HEK293 and CHO cells, with an individual RNA preparation (HEK293 and CHO) and duplicate ChIPs (CHO) derived from each of the three dishes. Transfection was performed in a single dish for NIH3T3 cells with three individual RNA preparations and triplicate ChIPs derived from the one dish.

44. Mohn, F., Weber, M., Schubeler, D. & Roloff, T. C. Methylated DNA immunoprecipitation (MeDIP). *Methods Mol. Biol.* **507**, 55–64 (2009).
45. Kireeva, M. L. *et al.* Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. *Mol. Cell* **30**, 557–566 (2008).
46. Awrey, D. E. *et al.* Transcription elongation through DNA arrest sites. A multistep process involving both RNA polymerase II subunit RPB9 and TFIIS. *J. Biol. Chem.* **272**, 14747–14754 (1997).
47. Kireeva, M. L., Lubkowska, L., Komissarova, N. & Kashlev, M. Assays and affinity purification of biotinylated and nonbiotinylated forms of double-tagged core RNA polymerase II from *Saccharomyces cerevisiae*. *Methods Enzymol.* **370**, 138–155 (2003).
48. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
49. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
50. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
51. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
52. Ramsköld, D., Wang, E. T., Burge, C. B. & Sandberg, R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput. Biol.* **5**, e1000598 (2009).
53. Wall, M. E., Dycck, P. A. & Brettin, T. S. SVDMAN—singular value decomposition analysis of microarray data. *Bioinformatics* **17**, 566–568 (2001).
54. Wall, M. E., Rechsteiner, A. & Rocha, L. M. in *A Practical Approach to Microarray Data Analysis* (eds Berrar, D. P., Dubitzky, W. & Granzow, M.) pp. 91–109 (Springer, 2003).

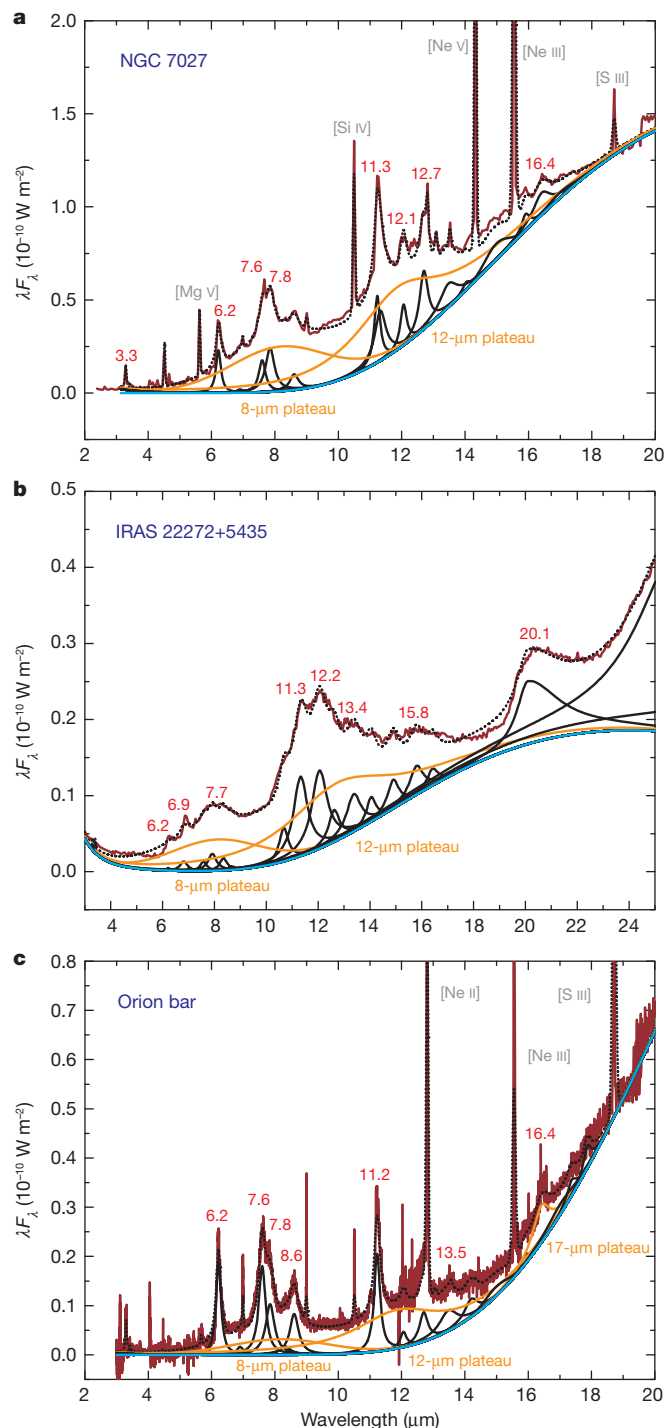
Mixed aromatic–aliphatic organic nanoparticles as carriers of unidentified infrared emission features

Sun Kwok¹ & Yong Zhang¹

Unidentified infrared emission bands at wavelengths of 3–20 micrometres are widely observed in a range of environments in our Galaxy and in others¹. Some features have been identified as the stretching and bending modes of aromatic compounds^{2,3}, and are commonly attributed to polycyclic aromatic hydrocarbon molecules^{4,5}. The central argument supporting this attribution is that single-photon excitation of the molecule can account for the unidentified infrared emission features observed in ‘cirrus’ clouds in the diffuse interstellar medium⁶. Of the more than 160 molecules identified in the circumstellar and interstellar environments, however, not one is a polycyclic aromatic hydrocarbon molecule. The detections of discrete and broad aliphatic spectral features suggest that the carrier of the unidentified infrared emission features cannot be a pure aromatic compound. Here we report an analysis of archival spectroscopic observations and demonstrate that the data are most consistent with the carriers being amorphous organic solids with a mixed aromatic–aliphatic structure. This structure is similar to that of the organic materials found in meteorites, as would be expected if the Solar System had inherited these organic materials from interstellar sources.

For the past 20 years, polycyclic aromatic hydrocarbon (PAH) molecules have commonly been considered the carriers of unidentified infrared emission (UIE) features. This hypothesis assumes that the UIE features are the result of infrared fluorescence from small (~50-carbon-atom) gas-phase PAH molecules being pumped by far-ultraviolet photons⁷. In spite of its popularity, the PAH hypothesis does not provide

Figure 1 | Mixed aromatic and aliphatic features in the infrared spectra of circumstellar and interstellar nebulae. **a–c**, Spectral decompositions of the UIE features of the planetary nebula NGC 7027 (**a**), the proto-planetary nebula IRAS 22272+5435 (**b**) and the Orion bar photodissociation region (**c**), showing a mix of aromatic, aliphatic and continuum features. The observed flux at wavelength λ , F_λ , is proportional to the emission intensity at that wavelength. A series of discrete features (black lines) and plateau features (orange lines; in **c**, the 17- μm plateau represents the 15–20- μm range) superposed on a continuum (blue line) have been fitted to the observed data. The UIE and plateau features (wavelengths in micrometres), as well as some of the atomic lines, are marked. The observed spectra are shown as solid red lines and the fitted spectra are shown as dotted black lines. The origin of the 20.1- μm feature in the IRAS 22272+5435 spectrum is currently unidentified. The spectral data for NGC 7027, IRAS 22272+5435 and the Orion bar are retrieved from the Infrared Space Observatory archive. For the spectral decomposition, we used the IDL package PAHFIT originally developed to fit the Spitzer Space Telescope Infrared Spectrograph spectra of nearby galaxies. The model spectra take into account the contributions from the stellar continuum, the thermal dust continuum, H₂ emission, atomic emission lines, the UIE features (both aromatic and aliphatic) and the plateau emission features. The optimal fitting to the observed spectra is achieved through the Levenberg–Marquardt least-squares algorithm. A modified blackbody model for the emission intensity at wavelength λ , $I_\lambda \propto \lambda^{-\alpha} B_\lambda(T)$, where $B_\lambda(T)$ is the blackbody function with a temperature T , is used to fit the continuum. The aromatic, aliphatic and plateau features are fitted with assumed Drude profiles $I_\lambda \propto \gamma^2 [(\lambda/\lambda_0 - \lambda_0/\lambda)^2 + \gamma^2]^{-1}$, where λ_0 is the central wavelength and γ is the fractional full-width at half-maximum of each feature.



¹Department of Physics, Faculty of Science, The University of Hong Kong, Pokfulam Road, Hong Kong, China.

a good explanation for the observed spectral behaviour. PAH molecules are fused ring molecules made up of carbon and hydrogen, and their vibrational bands are sharp and the peak wavelengths well defined. To fit the broad profiles of the UIE features seen in astronomical spectra, it is necessary to use a complex mixture of PAHs of different sizes, structures and charge states, and to utilize empirical feature profiles^{8,9}. Because PAH molecules require ultraviolet photons to excite them, they cannot explain the presence of UIE features in reflection nebulae¹⁰ and proto-planetary nebulae¹¹ where the central stars are cool and there is no ultraviolet background radiation. To account for these facts, the PAH model has to be revised to include large clusters and other ionization states.

The central argument for the PAH hypothesis is that single-photon excitation of PAH molecules can account for the 12- μm excess emission observed in cirrus clouds in the diffuse interstellar medium by the Infrared Astronomical Satellite (IRAS). However, the UIE-band flux ratios in the diffuse H II regions in the Carina nebula are nearly constant over a range of three orders of magnitude in background radiation¹². The shapes and peak wavelengths of the UIE features are independent of the temperature of the central stars providing the excitation¹⁰. Furthermore, PAH molecules have strong and narrow absorption features in the ultraviolet, but these are not observed in interstellar extinction curves¹³. However, the 3.4- μm aliphatic carbon-hydrogen stretching mode is commonly observed in absorption in the diffuse interstellar medium¹⁴. Although their rotational and vibrational frequencies are well known, not a single PAH molecule has yet been identified in space¹⁵.

Other arguments have been made to support the PAH hypothesis: the asymmetric profiles of the UIE features can be explained by anharmonicity associated with molecular emission, and the observed feature-to-continuum ratio is high and therefore implies that the carrier is a molecule⁷. Laboratory spectra of mixed aromatic and aliphatic solid materials have asymmetric profiles¹⁶, which can more

naturally explain the observations. Although the observed feature-to-continuum ratio is high in the diffuse interstellar medium, it is not high in the spectra of planetary and proto-planetary nebulae. Even in the diffuse interstellar medium, the strength of the UIE features are strongly correlated with the dust continuum, suggesting a possible physical relationship between the two components¹⁷.

The basic premise of the PAH hypothesis does not concern the chemical composition of the carrier so much as its size. As long as the carrier is a nanoparticle that can undergo transient heating, it will satisfy the excitation requirement. Laboratory experiments have yielded carbon nanoparticles with structures of sp^2 rings connected by networks of aliphatic chains^{18,19}, as well as fullerene fragments linked by aliphatic groups²⁰. These nanoparticles are likely to be constituents in circumstellar and interstellar environments. Furthermore, it has been proposed that the possible sudden release of chemical energy as a source of transient heating of small grains will allow much larger particles to radiate in the near-infrared, further weakening the PAH hypothesis²¹.

An alternative explanation to the UIE bands is that they are emitted by complex organic solids with disorganized structures. These solids intrinsically have broad emission profiles, and the features often sit on even broader emission plateaux several micrometres in width. It has been argued for some time that the observed spectral properties of UIE bands resemble those of coal and kerogen^{22,23}. Coal and kerogen are amorphous organic solids with a mixed sp^2 - sp^3 composition with randomly oriented aromatic ring units linked by long, aliphatic chains. Their mixed sp^2 - sp^3 chemistry gives rise to the discrete aromatic and aliphatic emission features and the broad plateau features¹⁶.

To provide a quantitative comparison between the two models, we have performed spectral decomposition of several sources with strong UIE features. Figure 1 shows a fit to the infrared spectra of the planetary nebula NGC 7027, the proto-planetary nebula IRAS 22272+5435 and

Table 1 | Strengths of the UIE discrete and plateau features

Discrete features* (%)										
	Aromatic					Aliphatic			Unknown	
λ (μm)	3.3	6.2	7.7	8.6	11.3	3.4	6.9	15.8	16.4	18.9
NGC 7027	0.32	1.2	2.8	0.58	3.1	0.07	0.11	0.31	0.84	0.0
IRAS 22272+5435†	0.08	0.05	0.30	0.11	3.76	0.15	0.43	0.94	0.41	—
Orion bar	0.67	3.8	7.0	2.2	2.6	0.13	0.37	0.24	2.0	0.0
V2361 Cygni	0.27‡	0.87§	0.60	0.03	—	—	0.25	—	—	—
V2362 Cygni	—	5.2	2.4	0.8	—	—	3.2	—	—	—
Plateau features (%)										
	Aromatic					Aliphatic			Unknown	
λ (μm)						8	12	17		
NGC 7027						18.8	17.2	0.33		
IRAS 22272+5435						12.5	18.6	—		
Orion bar						9.3	15.1	0.78		
V2361 Cygni						11.1	1.3	—		
V2362 Cygni						17.1	1.2	1.1		
Continuum										
	Percentage of total flux		Temperature (K)		α		Total flux (3–20 μm) (W m^{-2})			
NGC 7027	50.7		100		2		8.9 (–11)			
IRAS 22272+5435	40.8		100		2		2.5 (–11)			
Orion bar	47.3		70		2		2.4 (–11)			
V2361 Cygni	84.8		350		0.2		1.8 (–13)			
V2362 Cygni	66.8		365		0.5		2.1 (–13)			

The UIE phenomenon is complex. In addition to the commonly observed 3.3-, 6.2-, 7.7-, 8.6- and 11.3- μm aromatic features, there are also aliphatic features at 3.4 and 6.9 μm , arising respectively from symmetric and asymmetric carbon-hydrogen stretching and bending modes of methyl and methylene groups attached to aromatic rings. Features at 15.8, 16.4, 17.4 (not shown), 17.8 (not shown), and 18.9 μm have been found in proto-planetary nebulae¹¹, reflection nebulae¹⁰ and galaxies. In addition to the discrete features, broad emission features up to several micrometres in width are also seen. The 8- and 12- μm plateau features and a broad feature covering the 15–20- μm range (represented by 17 μm in the table) have been detected in young stellar objects, compact H II regions and planetary nebulae. The 8- and 12- μm plateaux are broad emission features (full-width at half-maximum, 2–4 μm) and can be identified as collective in-plane and, respectively, out-of-plane bending modes of a mixture of aliphatic side groups attached to aromatic rings²⁷. The 15–20- μm plateau feature is also found to be strong in some proto-planetary nebulae. This table summarizes the relative contributions of these components of the spectra shown in Figs 1 and 3.

*The total fluxes and percentages refer to the values emitted in the 3–20 μm range for NGC 7027 and Orion bar, 5–20 μm for V2362 Cygni and V2361 Cygni, and 3–25 μm for IRAS 22272+5435.

†For IRAS 22272+5435, there are additional features that contribute to the flux in the 3–25- μm range. Features at 12.2, 13.4 and 20.1 μm contribute 3.83, 1.53 and 2.56%, respectively. Some of the contributions also come from the broad 26- μm feature.

‡This entry refers to the spectral feature at 5.3 μm .

§This entry refers to the spectral feature at 6.3 μm .

||This entry refers to the spectral feature at 7.2 μm .

the Orion bar, a photodissociation region in the Orion nebula, using a set of discrete UIE features, broad plateau features, and the underlying dust continuum. The breakdown of contributions to the total fluxes from various components is summarized in Table 1. The strongest component is the continuum, which contributes approximately half of the total fluxes emitted in the 3–20- μm region. The next strongest are the plateau features, which account for 36, 31 and 25% of the total fluxes from NGC 7027, IRAS 22272 + 5435 and the Orion bar, respectively, compared with the totals of 8, 4 and 16% from the aromatic features. The aliphatic branches probably constitute a significant fraction of the material in each of the three sources.

The above fitting results show that the carrier of the UIE features includes a mixture of aromatic and aliphatic components, and is not a pure or predominantly aromatic compound. Because the carrier is formed from a mixture of cosmic gases, it is likely that the compound will include other abundant elements such as oxygen, nitrogen, sulphur and so on, in addition to carbon and hydrogen. These impurities may also have spectral signatures that can be identified by observations at higher spectral resolution. A sketch of the proposed chemical structure is shown in Fig. 2.

The best way to study the origin of the UIE features is to observe them when they are formed. From observations of objects in the late stages of stellar evolution, we know that the UIE features develop in the circumstellar environment within a few hundred years after the termination of the asymptotic giant branch²⁴. Spectroscopic observations of novae have shown that the 3.3- and 3.4- μm features appear soon after dust condensation²⁵. Theoretically, it is difficult to understand how complex organics can form under such low-density conditions, but novae are observed to change from a pure gas spectrum to a dust-dominated spectrum over the course of days²⁶. In Fig. 3, we show a fit to Spitzer spectra of the novae V2362 Cygni and V2361 Cygni. It is expected that a mixture of miscellaneous aliphatic branches will attach to the newly formed ring clusters. The prominence of the plateau features reflects this early stage of organic dust condensation²⁷.

We note that the dominant organic content in carbonaceous chondrites is a kerogen-like macromolecular solid referred to as insoluble organic matter. Recent laboratory analysis of the insoluble organic matter in the Murchison meteorite has suggested that it has a chemical structure very similar to that which we propose here^{28,29}. The presence of insoluble organic matter in meteorites is evidence that complex

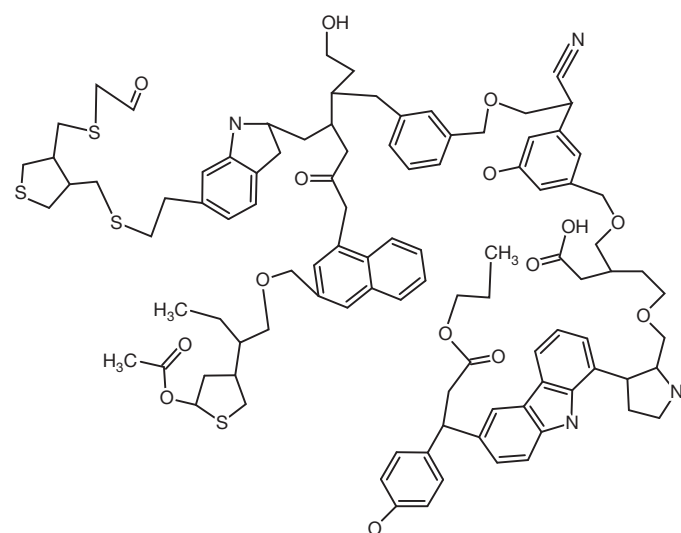


Figure 2 | Proposed structure of the carrier of UIE features. The structure is characterized by a highly disorganized arrangement of small units of aromatic rings linked by different kinds of aliphatic chain. Other impurities such as oxygen, nitrogen and sulphur are also commonly present. This structure contains about 100 carbon atoms and a typical nanoparticle may consist of multiple structures similar to this one.

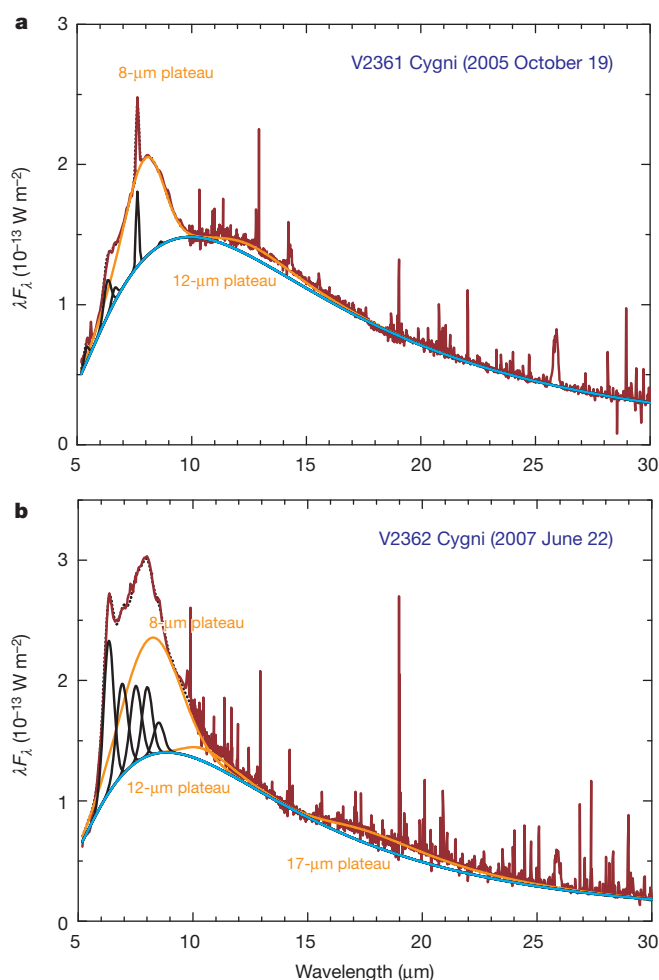


Figure 3 | Emergence of complex organics after nova outburst. **a, b,** Fits to the Spitzer Infrared Spectrograph spectra of novae V2361 Cygni (**a**) and V2362 Cygni (**b**) 251 and 446 days after their respective outbursts. In addition to the gas emission line spectrum, both spectra have developed strong dust continua, and the 8- and 12- μm plateau features are clearly present. The continua of V2361 Cygni and V2362 Cygni are fitted by modified blackbody intensities of the respective forms $\lambda^{-0.2}B_{\lambda}(350\text{ K})$ and $\lambda^{-0.5}B_{\lambda}(365\text{ K})$ (blue lines). The orange lines are the 8-, 12- and 17- μm plateau features and the solid black lines are discrete features at 5.3, 6.3, 6.9, 7.2 and 8.6 μm for V2361 Cygni and at 6.2, 6.9, 7.6, 7.8 and 8.6 μm for V2362 Cygni. The observed spectra are shown as solid red lines and the fitted spectra are shown as dotted black lines. The presence of the 8- and 12- μm plateau features suggests that the aliphatic component is the first to emerge after dust condensation. Because of the large number of emission lines, the atomic lines are not included in the fitting.

organic solids form in nature with no difficulty. The fact that insoluble organic matter and circumstellar dust have similar chemical structures offers the possibility that Solar System organics may have a stellar connection.

Received 1 April; accepted 30 August 2011.

Published online 26 October 2011.

1. Kwok, S. *Organic Matter in the Universe* (Wiley, 2011).
2. Knacke, R. F. Carbonaceous compounds in interstellar dust. *Nature* **269**, 132–134 (1977).
3. Duley, W. W. & Williams, D. A. The infrared spectrum of interstellar dust: surface functional groups on carbon. *Mon. Not. R. Astron. Soc.* **196**, 269–274 (1981).
4. Allamandola, L. J., Tielens, A. G. G. M. & Barker, J. R. Interstellar polycyclic aromatic hydrocarbons: the infrared emission bands, the excitation/emission mechanism and the astrophysical implications. *Astrophys. J. Suppl. Ser.* **71**, 733–775 (1989).
5. Puget, J. L. & Léger, A. A new component of the interstellar matter: small grains and large aromatic molecules. *Annu. Rev. Astron. Astrophys.* **27**, 161–198 (1989).

6. Sellgren, K. The near-infrared continuum emission of visual reflection nebulae. *Astrophys. J.* **277**, 623–633 (1984).
7. Tielens, A. G. G. M. Interstellar polycyclic aromatic hydrocarbon molecules. *Annu. Rev. Astron. Astrophys.* **46**, 289–337 (2008).
8. Peeters, E. *et al.* The rich 6 to 9 μm spectrum of interstellar PAHs. *Astron. Astrophys.* **390**, 1089–1113 (2002).
9. Draine, B. T. & Li, A. Infrared emission from interstellar dust. IV. The silicate-graphite-PAH model in the post-Spitzer era. *Astrophys. J.* **657**, 810–837 (2007).
10. Uchida, K. I., Sellgren, K., Werner, M. W. & Houdashelt, M. L. Infrared Space Observatory mid-infrared spectra of reflection nebulae. *Astrophys. J.* **530**, 817–833 (2000).
11. Kwok, S., Volk, K. & Hrivnak, B. J. Chemical evolution of carbonaceous materials in the last stages of stellar evolution. *Astron. Astrophys.* **350**, L35–L38 (1999).
12. Onaka, T. Interstellar dust: what do space observations tell us? *Adv. Space Res.* **25**, 2167–2176 (2000).
13. Clayton, G. C. *et al.* The role of polycyclic aromatic hydrocarbons in ultraviolet extinction. I. Probing small molecular polycyclic aromatic hydrocarbons. *Astrophys. J.* **592**, 947–952 (2003).
14. Pendleton, Y. J. & Allamandola, L. J. The organic refractory material in the diffuse interstellar medium: mid-infrared spectroscopic constraints. *Astrophys. J. Suppl. Ser.* **138**, 75–98 (2002).
15. Pilleri, P. *et al.* Search for corannulene ($\text{C}_{20}\text{H}_{10}$) in the Red Rectangle. *Mon. Not. R. Astron. Soc.* **397**, 1053–1060 (2009).
16. Guillois, O., Nenner, I., Papoular, R. & Reynaud, C. Coal models for the infrared emission spectra of proto-planetary nebulae. *Astrophys. J.* **464**, 810–817 (1996).
17. Kahanpää, J., Mattila, K., Lehtinen, K., Leinert, C. & Lemke, D. Unidentified infrared bands in the interstellar medium across the Galaxy. *Astron. Astrophys.* **405**, 999–1012 (2003).
18. Sakata, A., Wada, S., Onaka, T. & Tokunaga, A. T. Infrared spectrum of quenched carbonaceous composite (QCC). II. A new identification of the 7.7 and 8.6 micron unidentified infrared emission bands. *Astrophys. J.* **320**, L63–L67 (1987).
19. Duley, W. W. & Hu, A. Polyyenes and interstellar carbon nanoparticles. *Astrophys. J.* **698**, 808–811 (2009).
20. Jäger, C., Huisken, F., Mutschke, H., Jansa, I. L. & Henning, T. H. Formation of polycyclic aromatic hydrocarbons and carbonaceous solids in gas-phase condensation experiments. *Astrophys. J.* **696**, 706–712 (2009).
21. Duley, W. W. & Williams, D. A. Excitation of the aromatic infrared emission bands: chemical energy in hydrogenated amorphous carbon particles? *Astrophys. J.* **737**, L44 (2011).
22. Papoular, R., Conrad, J., Giuliano, M., Kister, J. & Mille, G. A coal model for the carriers of the unidentified IR bands. *Astron. Astrophys.* **217**, 204–208 (1989).
23. Papoular, R. The use of kerogen data in understanding the properties and evolution of interstellar carbonaceous dust. *Astron. Astrophys.* **378**, 597–607 (2001).
24. Kwok, S. The synthesis of organic and inorganic compounds in evolved stars. *Nature* **430**, 985–991 (2004).
25. Evans, A. *et al.* Infrared spectroscopy of nova Cassiopeiae 1993. IV. A closer look at the dust. *Mon. Not. R. Astron. Soc.* **360**, 1483–1492 (2005).
26. Ney, E. P. & Hatfield, B. F. The isothermal dust condensation of Nova Vulpeculae 1976. *Astrophys. J.* **219**, L111–L115 (1978).
27. Kwok, S., Volk, K. & Bernath, P. On the origin of infrared plateau features in proto-planetary nebulae. *Astrophys. J.* **554**, L87–L90 (2001).
28. Derenne, S. & Robert, F. Model of molecular structure of the insoluble organic matter isolated from Murchison meteorite. *Meteorit. Planet. Sci.* **45**, 1461–1475 (2010).
29. Cody, G. D. *et al.* Establishing a molecular relationship between chondritic and cometary organic solids. *Proc. Natl Acad. Sci. USA* advance online publication, (<http://dx.doi.org/10.1073/pnas.1015913108>) (4 April 2011).
30. Sellgren, K., Uchida, K. I. & Werner, M. W. The 15–20 μm Spitzer spectra of interstellar emission features in NGC 7023. *Astrophys. J.* **659**, 1338–1351 (2007).

Acknowledgements We thank A. Tang for technical assistance in the preparation of this manuscript. This work was supported by a grant to S.K. from the Research Grants Council of the Hong Kong Special Administrative Region, China (project no. HKU 7027/11P).

Author Contributions S.K. designed the research and wrote the paper. Y.Z. performed data analysis and model fitting.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.K. (sunkwok@hku.hk).

Room temperature coherent control of defect spin qubits in silicon carbide

William F. Koehl¹, Bob B. Buckley¹, F. Joseph Heremans¹, Greg Calusine¹ & David D. Awschalom¹

Electronic spins in semiconductors have been used extensively to explore the limits of external control over quantum mechanical phenomena¹. A long-standing goal of this research has been to identify or develop robust quantum systems that can be easily manipulated, for future use in advanced information and communication technologies². Recently, a point defect in diamond known as the nitrogen–vacancy centre has attracted a great deal of interest because it possesses an atomic-scale electronic spin state that can be used as an individually addressable, solid-state quantum bit (qubit), even at room temperature³. These exceptional quantum properties have motivated efforts to identify similar defects in other semiconductors, as they may offer an expanded range of functionality not available to the diamond nitrogen–vacancy centre⁴. Notably, several defects in silicon carbide (SiC) have been suggested as good candidates for exploration, owing to a combination of computational predictions and magnetic resonance data^{4–10}. Here we demonstrate that several defect spin states in the 4H polytype of SiC (4H-SiC) can be optically addressed and coherently controlled in the time domain at temperatures ranging from 20 to 300 kelvin. Using optical and microwave techniques similar to those used with diamond nitrogen–vacancy qubits, we study the spin-1 ground state of each of four inequivalent forms of the neutral carbon–silicon divacancy, as well as a pair of defect spin states of unidentified origin. These defects are optically active near telecommunication wavelengths¹¹, and are found in a host material for which there already exist industrial-scale crystal growth¹² and advanced microfabrication techniques¹³. In addition, they possess desirable spin coherence properties that are comparable to those of the diamond nitrogen–vacancy centre. This makes them promising candidates for various photonic, spintronic and quantum information applications that merge quantum degrees of freedom with classical electronic and optical technologies^{2,14–17}.

SiC is a wide-bandgap, group IV semiconductor with well-established growth and device engineering protocols. Inch-scale, high-quality bulk and epitaxial single crystals of SiC are commercially available¹², and a wide variety of advanced electronic^{13,18,19}, optoelectronic^{13,20} and electromechanical²⁰ devices have been successfully fabricated from SiC. Other notable features of SiC are the existence of many polytypes, each with a different bandgap¹⁸; a native oxide that allows for sophisticated complementary metal-oxide-semiconductor circuitry^{18,19}; and its common use as a growth substrate for other technologically relevant materials, such as gallium nitride and graphene^{21,22}. Because of this versatility, the existence of SiC-based analogues to the diamond nitrogen–vacancy centre could lead to exciting developments in quantum engineering and the study of quantum phenomena in the solid state.

Here we are interested in defects that possess a tightly bound electronic spin that can be polarized and measured through the absorption and luminescence of light. In all experiments, samples diced from a 2-inch wafer of high-purity semi-insulating (HPSI) 4H-SiC are optically excited with below-bandgap light from an 853 nm (1.45 eV) diode laser. Photoluminescence spectra taken at sample temperatures ranging from 20 to 300 K are shown in Fig. 1a. At 20 K, several sharp lines with

energies ranging from 1.09 to 1.20 eV are apparent, along with a series of broader features found from 0.9 to 1.20 eV. At low temperatures, luminescence from a single species of optically active defect commonly consists of a sharp zero-phonon line (ZPL) and a broad sideband composed of lower-energy phonon replicas of the ZPL. If several distinct species are probed simultaneously, the total photoluminescence signal will be the sum of its parts, so that the ZPL of one defect species may lie on top of the phonon sideband of another. This is what is observed in Fig. 1a. An expanded view of the 20 K photoluminescence from 1.09 to 1.20 eV can be seen in Fig. 1b, where six sharp lines have been given the labels PL1–PL6. Four of these, PL1–PL4, form a group of lines known in previous literature by the singular label, UD-2 (ref. 23). The other two, PL5 and PL6, are close to one another in energy and share several other features that are described in the Supplementary Information. This suggests that PL5 and PL6 are related, but their origins are currently unidentified. At higher temperatures, all six lines simultaneously broaden and decrease in height, so that they are almost completely indistinguishable from the phonon sidebands by 200 K. However, luminescence through these phonon sidebands is still clearly visible, even at 300 K.

Previous results from photo-enhanced electron spin resonance and annealing experiments have indicated that the four UD-2 luminescence lines (PL1–PL4) originate from four distinct forms of the neutral divacancy, which is an uncharged defect complex consisting of a carbon vacancy adjacent to a silicon vacancy^{24,25}, $[V_{Si}-V_C]^0$. Two forms of the divacancy, which are often given the labels (*hh*) and (*kk*), are oriented along the *c* axis of the crystal. The other two, (*hk*) and (*kh*), are oriented along the basal bond directions. These four defects give rise to the *P6b/P6'b* (*c*-axis) and *P7b/P7'b* (basal) spin-resonance signals that are often observed in electron spin resonance studies of 4H-SiC, and that have been shown to correspond to spin-1 ground states that can be spin-polarized with incident light^{6,7,26}. As discussed in the Supplementary Information, we use angle-resolved magnetoluminescence and continuous-wave optically detected magnetic resonance techniques to confirm that each of the four UD-2 luminescence lines corresponds to a different form of the neutral divacancy. These experiments also show that the photoluminescence intensity of each divacancy is modulated by the dynamics of its ground state spin.

An abridged version of these data can be seen in Fig. 1c–h. In each panel, the normalized change in photoluminescence ($\Delta PL/PL$) from one of the six lines, PL1–PL6, is seen as a function of applied microwave frequency at 20 K and zero magnetic field. The direction of the microwave driving field varied across the observed spin ensemble, so that spins oriented in any direction could be driven. Spin resonances are observed in all six lines. In Fig. 1c we see a sharp decrease in PL1 luminescence resulting from microwave-induced transitions between sublevels of the (*hh*) *c*-axis divacancy ground state spin. The central frequency of this resonance corresponds to the axially symmetric spin splitting (*D*) of this defect spin state. In Fig. 1d, there is a similarly shaped increase in PL2 luminescence centred at a frequency corresponding to *D* for the other form of the *c*-axis divacancy, (*kk*). A pair of resonances is observed in both Fig. 1e and f, which show $\Delta PL/PL$ for

¹Center for Spintronics and Quantum Computation, University of California, Santa Barbara, California 93106, USA.

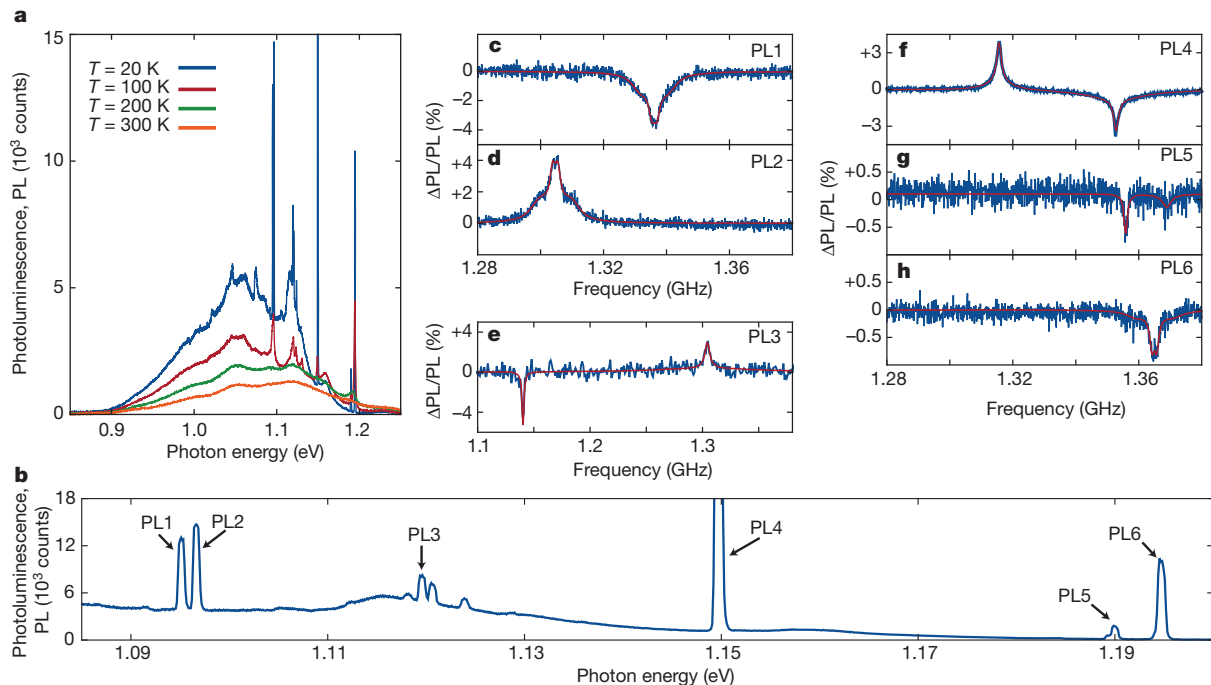


Figure 1 | Optical detection of defect spins in 4H-SiC. **a**, Photoluminescence spectra of HPSI 4H-SiC at sample temperatures ranging from 20 to 300 K. **b**, An expanded view of low-temperature (20 K) photoluminescence showing the six defect lines (PL1–PL6) discussed in the text. PL1–PL4 are the four distinct

forms of the neutral divacancy, whereas PL5 and PL6 remain unidentified. **c–h**, The normalized change in luminescence ($\Delta\text{PL}/\text{PL}$) collected from PL1 (**c**), PL2 (**d**), PL3 (**e**), PL4 (**f**), PL5 (**g**) or PL6 (**h**) as a function of applied microwave frequency at 20 K. Solid red lines in **c–h** are fits.

PL3 and PL4, respectively. Each figure corresponds to one of the two basal forms of the divacancy. Owing to the hexagonal crystal structure of 4H-SiC, these basal defects possess a lower symmetry than their *c*-axis counterparts. Their ground state spins therefore exhibit a transverse anisotropy spin splitting (E) in addition to D . This results in two zero-field resonances at frequencies defined by $(D \pm E)$, as seen in the data⁷. Figure 1g and h shows spin resonances observed in PL5 and PL6, which are emitted by the basal and *c*-axis forms of the unidentified defects, respectively.

As shown in the Supplementary Information, the resonances in Fig. 1c–h do not respond uniformly to elevated temperatures. The basal divacancy resonances are no longer detectable at 100 K, whereas the (*kk*) and (*hh*) *c*-axis divacancy resonances are observed up to temperatures of 200 and 250 K, respectively. The resonances of the unidentified defects persist to room temperature however, and decrease in magnitude by only about 30% between 20 and 300 K.

We now demonstrate coherent control and direct optical measurement of each of these six defect spins by extending our measurements to the time domain. The time-resolved measurements that we describe include three steps. First, we polarize the defect spin ensemble with a pulse of light. Then, we coherently manipulate the ensemble with pulsed microwaves. Finally, we excite the ensemble with a second pulse of light and measure its photoluminescence intensity, which is spin-dependent. For instance, the results of such an experiment demonstrating coherent control over the PL4 basal divacancy spins at 20 K are shown in Fig. 2a. We coherently rotate these spins by applying a pulse of 1.3526 GHz microwaves resonant with the high-frequency transition seen in Fig. 1f. By varying the duration of the microwave pulse, we observe Rabi oscillations. The magnetic field component of the microwave driving field was directed only along the *c* axis of the material, and no static magnetic field was applied. This measurement was made at several different microwave powers, and the resulting spin dynamics were fitted to an exponentially decaying cosine to determine the Rabi frequency, ω_R . As can be seen from the data in Fig. 2b, the Rabi frequency increases linearly with the square root of microwave power, as expected from the Rabi formula²⁷.

A Ramsey pulse sequence²⁸ is used to observe free induction decay of the same defect ensemble at 20 K (Fig. 2c), and the resulting dynamics are fitted to reveal an inhomogeneous spin coherence time

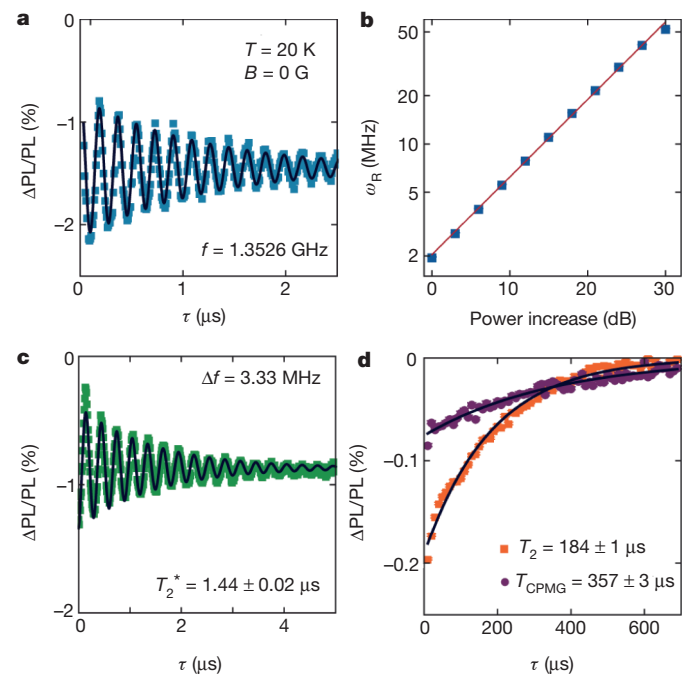


Figure 2 | Time-resolved dynamics of basal divacancy spins at 20 K. **a**, Rabi oscillations observed in basal divacancy spins at 20 K and 0 G. **b**, Rabi frequency (ω_R) measured as a function of relative microwave power. In both **a** and **b**, the driving frequency ($f = 1.3526$ GHz) was resonant with the high-frequency feature seen in Fig. 1f. **c**, Ramsey decay observed in the same population of basal divacancy spins, revealing that $T_2^* = 1.44 \mu\text{s}$. Oscillations are due to a microwave detuning of $\Delta f = 3.33$ MHz. **d**, Optically detected Hahn echo and CPMG decays, with $T_2 = 184 \mu\text{s}$ and $T_{\text{CPMG}} = 357 \mu\text{s}$. Solid lines are fits; error bars indicate ± 1 s.d.

of $T_2^* = 1.44 \pm 0.02 \mu\text{s}$ for the ground state spins. The Ramsey decay oscillates because the driving field was detuned from resonance by 3.33 MHz. Data from Hahn echo and $3\text{-}\pi$ pulse Carr-Purcell-Meiboom-Gill (CPMG) pulse sequences²⁸ are shown in Fig. 2d, revealing homogenous spin coherence times of $T_2 = 184 \pm 1 \mu\text{s}$ and $T_{\text{CPMG}} = 357 \pm 3 \mu\text{s}$ at 20 K in our samples. This is the same order of magnitude as T_2 for diamond nitrogen-vacancy centre ensembles surrounded by a spin bath composed of ^{13}C nuclear spins and background paramagnetic impurities²⁹.

Using the same c -axis driving field geometry, Rabi oscillations were also observed at the high-frequency resonance of PL3. However, to efficiently couple to the low-frequency resonances of PL3 or PL4, a driving field with a magnetic field component oriented perpendicular to the c axis needed to be used. The reduced symmetry of a basal divacancy leads to different microwave coupling geometries for the two observed spin transitions. The data therefore suggest that the low-frequency dipole transitions of the basal divacancies are perpendicular to the c axis whereas the high-frequency dipole transitions are not.

The c -axis forms of the neutral divacancy offer a more direct comparison to the diamond nitrogen-vacancy centre because they share the same symmetry and are predicted to have a similar electronic structure⁵. Rabi and Ramsey oscillations induced by an in-plane driving field are observed for the (hh) divacancy at 200 K and 52 G (Fig. 3a and b), illustrating coherent control over this defect. The 52 G field was applied along the defect axis in order to split the $m_s = \pm 1$ spin sub-levels of the ground state, which are roughly degenerate at zero magnetic field. Similar results were also observed for the (kk) c -axis divacancy at 200 K (see Supplementary Information). While we find that $T_2^* = 182 \pm 6 \text{ ns}$, a Hahn echo measurement taken at 113 G (Fig. 3c) shows that $T_2 = 263 \pm 4 \mu\text{s}$. Additionally, periodic modulations of the Hahn echo envelope are observed in Fig. 3c that appear when a magnetic field is applied along the c axis of the crystal. These modulations increase in frequency as the magnetic field is increased,

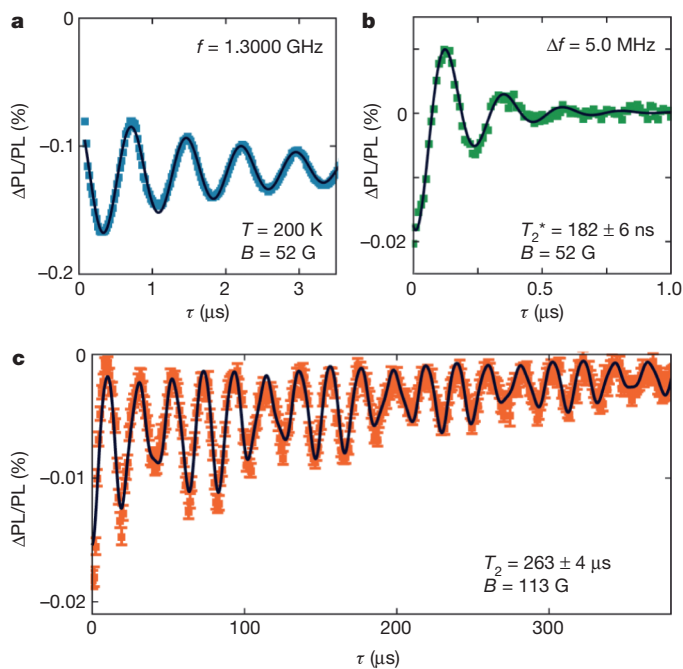


Figure 3 | Time-resolved dynamics of c -axis divacancy spins at 200 K. **a**, Rabi oscillations observed in c -axis (hh) divacancy spins at 200 K and 52 G. The magnetic field was oriented along the c axis. The driving frequency ($f = 1.3000 \text{ GHz}$) was resonant with the central frequency of the feature seen in Fig. 1c, once Zeeman shifted by 52 G. **b**, Ramsey decay observed in the same population of (hh) divacancy spins, revealing that $T_2^* = 182 \text{ ns}$. Oscillations are due to a microwave detuning of $\Delta f = 5.0 \text{ MHz}$. **c**, Hahn echo at 113 G, showing oscillations induced by the nuclear spin bath; here $T_2 = 263 \mu\text{s}$. Solid lines are fits; error bars indicate $\pm 1 \text{ s.d.}$

and fits to the data at a given field reveal that they contain two frequency components equal to the Larmor frequencies of ^{29}Si and ^{13}C nuclei. These modulations therefore represent coherent interactions between the (hh) divacancy spins and the surrounding nuclear spin bath³⁰.

Lastly, we demonstrate room temperature coherent control of the unidentified defects corresponding to PL5 and PL6. In Fig. 4a and b, we show Rabi measurements of the c -axis (PL6) and basal (PL5) orientations of this defect, respectively. These measurements were made at 300 K and 0 G, with an in-plane driving field along the $\langle 1100 \rangle$ crystal direction. Although the c -axis defect spins oscillate at one frequency, the basal defect spins do not. This beating effect is observed because the basal defect spins exist in three orientations that couple unequally to the in-plane driving field. Ramsey measurements on both the c -axis and basal orientations of this defect are shown in the Supplementary Information, and exhibit maximum T_2^* times of $214 \pm 3 \text{ ns}$ and $1,248 \pm 45 \text{ ns}$ respectively. Previous work using conventional electron spin resonance techniques has shown that silicon vacancies in SiC also exhibit room temperature Rabi oscillations⁸.

Hahn spin echo measurements at zero magnetic field reveal that the c -axis and basal orientations of the unidentified defect decay

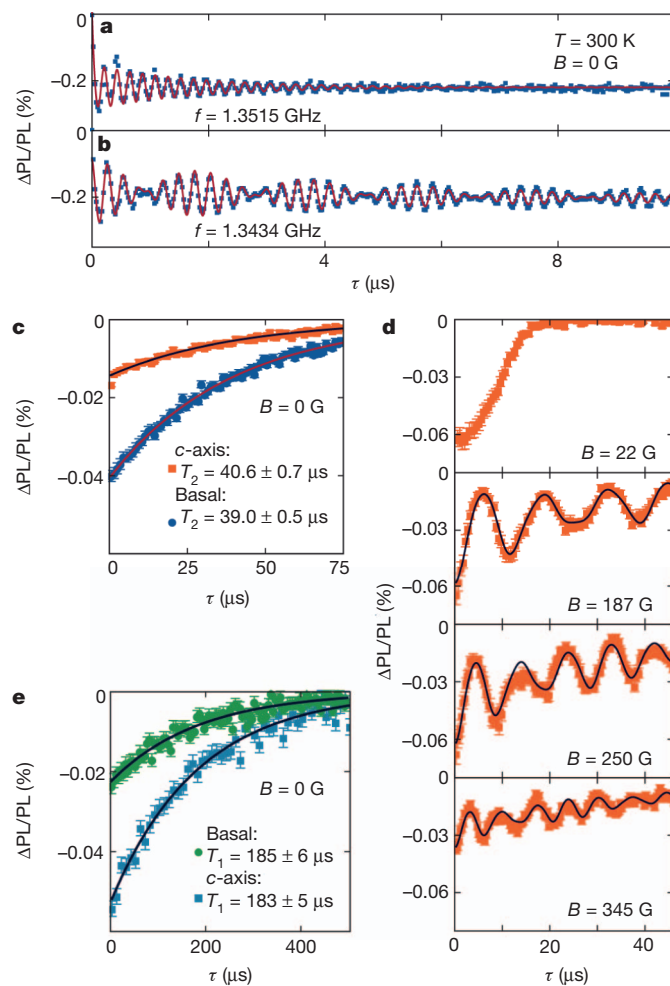


Figure 4 | Coherent control of defect spins in SiC at room temperature. **a**, **b**, Rabi oscillations observed at 300 K and 0 G in the c -axis (**a**) and basal (**b**) orientations of the unidentified defects, which emit PL6 and PL5, respectively. **c**, Hahn echo decays at 0 G for the same unidentified defects, revealing T_2 times of $\sim 40 \mu\text{s}$. **d**, Magnetic field dependence of the c -axis Hahn echo, showing oscillations induced by the nuclear spin bath. **e**, T_1 measurements for the c -axis and basal orientations, revealing long longitudinal spin relaxation times of $T_1 = 183 \mu\text{s}$ and $T_1 = 185 \mu\text{s}$ respectively. Solid lines are fits; error bars indicate $\pm 1 \text{ s.d.}$

exponentially with T_2 times of $40.6 \pm 0.7 \mu\text{s}$ and $39.0 \pm 0.5 \mu\text{s}$, respectively (Fig. 4c). As with the (hh) divacancy, the spin echo envelope of the c -axis unidentified defect becomes modulated when a non-zero magnetic field is applied along the c -axis of the crystal. At each field (Fig. 4d, bottom three panels), fits to the data again reveal two frequency components equal to the Larmor frequencies of ^{29}Si and ^{13}C nuclei. In Fig. 4e we measure the longitudinal spin relaxation time, T_1 , and find values of $183 \pm 5 \mu\text{s}$ and $185 \pm 6 \mu\text{s}$ for the c -axis and basal spins, respectively. These are relatively long T_1 times, suggesting that we are probing spins in a ground or metastable electronic state of the defect, rather than those of an orbital excited state with a short optical lifetime.

These experiments demonstrate that several highly controllable defect-based quantum states exist in SiC, and that they can be manipulated with the same techniques developed and optimized for use with diamond nitrogen–vacancy qubits. Notably, we have shown that two of these defects are capable of room temperature operation, and that differences in defect composition or orientation made possible by the binary nature of SiC lead to a rich assortment of electronic and spin properties. Because well-established growth and processing protocols already exist for SiC, these defects could be the basis for a previously unachievable degree of integration between defect qubits and existing device technologies.

METHODS SUMMARY

Samples were mounted in a liquid helium flow cryostat with optical and microwave access. A closed-loop sample heater was used to hold sample temperatures constant between 20 and 300 K. 853 nm diode laser light was used for sample excitation, and was pulsed using an acousto-optic modulator (AOM). A 14 mm focal length lens (Figs 1, 3 and 4) or a 60 \times microscope objective (Fig. 2) focused the light on the sample surface with spot diameters of approximately 15 or 1 μm , respectively. The laser power was ~ 20 mW at the sample surface. Two infrared detectors were used to measure photoluminescence: a liquid-nitrogen-cooled spectrometer (Fig. 1), or a femtowatt photoreceiver with an analogue voltage output (Figs 2–4). Analogue voltages were measured and processed digitally using a data acquisition system (DAQ). Both detectors had an 800–1,700 nm operating range. Microwaves were supplied to the cryostat by a signal generator and were modulated using a selection of microwave switches and mixers. A permanent magnet was used to apply a magnetic field along the c axis of the material in Figs 3 and 4d. Otherwise, no magnet was present.

Samples were diced from a 365- μm -thick wafer of HPSI 4H-SiC purchased from CREE (part number: W4TRD0R-0200), and had dimensions of roughly 2 mm \times 3 mm. Ring-shaped microwave sources 34 μm (Fig. 1) and 1 mm (Fig. 2) in diameter were patterned on the surfaces of two samples using standard photolithographic techniques. In both cases, a 10/90 nm Ti/Pt metallization was used. A third sample was mounted on top of a 1.8 mm wide by 9 mm long microwave stripline (Figs 3 and 4) made from RT/duroid 6002 (Rogers Corporation) plated with 1.8 μm of Au. All microwave devices were connected via wire bonds to the microwave line in the cryostat.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 July; accepted 19 September 2011.

1. Hanson, R. & Awschalom, D. D. Coherent manipulation of single spins in semiconductors. *Nature* **453**, 1043–1049 (2008).
2. Awschalom, D. D. & Flatté, M. E. Challenges for semiconductor spintronics. *Nature Phys.* **3**, 153–159 (2007).
3. Jelezko, F., Gaebel, T., Popa, I., Gruber, A. & Wrachtrup, J. Observation of coherent oscillations in a single electron spin. *Phys. Rev. Lett.* **92**, 076401 (2004).

4. Weber, J. R. *et al.* Quantum computing with defects. *Proc. Natl Acad. Sci. USA* **107**, 8513–8518 (2010).
5. Gali, A. Time-dependent density functional study on the excitation spectrum of point defects in semiconductors. *Phys. Status Solidi B* **248**, 1337–1346 (2011).
6. Baranov, P. G. *et al.* EPR identification of the triplet ground state and photoinduced population inversion for a Si-C divacancy in silicon carbide. *JETP Lett.* **82**, 441–443 (2005).
7. Son, N. T. *et al.* Divacancy in 4H-SiC. *Phys. Rev. Lett.* **96**, 055501 (2006).
8. Mizuochi, N. *et al.* Continuous-wave and pulsed EPR study of the negatively charged silicon vacancy with $S=3/2$ and C_{3v} symmetry in n -type 4H-SiC. *Phys. Rev. B* **66**, 235202 (2002).
9. Son, N. T., Zolnai, Z. & Janzén, E. Silicon vacancy related T_{V2a} center in 4H-SiC. *Phys. Rev. B* **68**, 205211 (2003).
10. Baranov, P. G., Bundakova, A. P. & Soltamov, A. A. Silicon vacancy in SiC as a promising quantum system for single-defect and single-photon spectroscopy. *Phys. Rev. B* **83**, 125203 (2011).
11. Saleh, B. E. A. & Teich, M. C. *Fundamentals of Photonics* Ch. 22 (Wiley, 1991).
12. Powell, A. *et al.* Growth of SiC substrates. *Int. J. High Speed Electron. Syst.* **16**, 751–777 (2006).
13. Zetterling, C.-M. (ed.) *Process Technology for Silicon Carbide Devices* (Institution of Electrical Engineers, 2002).
14. O'Brien, J. L., Furusawa, A. & Vučković, J. Photonic quantum technologies. *Nature Photon.* **3**, 687–695 (2009).
15. Maze, J. R. *et al.* Nanoscale magnetic sensing with an individual electronic spin in diamond. *Nature* **455**, 644–647 (2008).
16. Balasubramanian, G. *et al.* Nanoscale imaging magnetometry with diamond spins under ambient conditions. *Nature* **455**, 648–651 (2008).
17. Ladd, T. D. *et al.* Quantum computers. *Nature* **464**, 45–53 (2010).
18. Sadow, S. E. & Agarwal, A. (eds) *Advances in Silicon Carbide Processing and Applications* (Artech House, 2004).
19. Ryu, S.-H., Kornegay, K. T., Cooper, J. A. & Melloch, M. R. Digital CMOS IC's in 6H-SiC operating on a 5-V power supply. *IEEE Trans. Electron Dev.* **45**, 45–53 (1998).
20. Cheung, R. (ed.) *Silicon Carbide Microelectromechanical Systems for Harsh Environments* (Imperial College Press, 2004).
21. Liu, L. & Edgar, J. H. Substrates for gallium nitride epitaxy. *Mater. Sci. Eng. R* **37**, 61–127 (2002).
22. Berger, C. *et al.* Electronic confinement and coherence in patterned epitaxial graphene. *Science* **312**, 1191–1196 (2006).
23. Magnusson, B. & Janzén, E. Optical characterization of deep level defects in SiC. *Mater. Sci. Forum* **483–485**, 341–346 (2005).
24. Carlos, W. E., Glaser, E. R. & Shanabrook, B. V. Optical and magnetic resonance signatures of deep levels in semi-insulating 4H SiC. *Physica B* **340–342**, 151–155 (2003).
25. Carlos, W. E., Graces, N. Y., Glaser, E. R. & Fanton, M. A. Annealing of multivacancy defects in 4H-SiC. *Phys. Rev. B* **74**, 235201 (2006).
26. Son, N. T. *et al.* Identification of divacancies in 4H-SiC. *Physica B* **376–377**, 334–337 (2006).
27. Hanson, R., Gywat, O. & Awschalom, D. D. Room-temperature manipulation and decoherence of a single spin in diamond. *Phys. Rev. B* **74**, 161203(R) (2006).
28. de Lange, G., Wang, Z. H., Ristè, D., Dobrovitski, V. V. & Hanson, R. Universal dynamical decoupling of a single solid-state spin from a spin bath. *Science* **330**, 60–63 (2010).
29. Stanwix, P. L. *et al.* Coherence of nitrogen-vacancy electronic spin ensembles in diamond. *Phys. Rev. B* **82**, 201201(R) (2010).
30. van Oort, E. & Glasbeek, M. Optically detected low field electron spin echo envelope modulations of fluorescent N-V centers in diamond. *Chem. Phys.* **143**, 131–140 (1990).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to G. D. Fuchs, A. Janotti, D. M. Toyli, C. G. Van de Walle, J. B. Varley and J. R. Weber for discussions. We thank M. E. Nowakowski for help with sample preparation. This work was supported by the Air Force Office of Scientific Research (AFOSR) and the Defense Advanced Research Projects Agency (DARPA).

Author Contributions All authors helped to design the research, perform the research and write the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.D.A. (awsch@physics.ucsb.edu).

METHODS

Experimental set-up. In all experiments, samples were held at temperatures ranging from 20 to 300 K by mounting them in a liquid helium flow cryostat outfitted with a closed-loop sample heater. A window and a microwave feedthrough provided optical and microwave access to the sample. Light used for sample excitation came from an 853 nm diode laser that was first passed through an acousto-optic modulator (AOM) and then through a series of polarization optics before being reflected off a dichroic mirror designed to transmit wavelengths in the 932–1,300 nm range. The reflected laser beam was then focused by either a 14 mm focal length lens (Figs 1, 3 and 4) or a 60 \times microscope objective (Fig. 2) onto the sample surface with a spot diameter of roughly 15 or 1 μ m, respectively. The polarization optics were tuned to ensure that incident light at the sample surface was linearly polarized parallel to the table surface. A laser power of \sim 20 mW at the sample surface was used. Photoluminescence was collected through the same focusing optic and then transmitted through the dichroic and a 900 nm longpass filter to one of two infrared detectors, depending on the data type being collected. The first detector was a liquid-nitrogen-cooled spectrometer (Fig. 1); the second was a femtowatt photoreceiver with a 20 Hz bandwidth and an analogue voltage output (Figs 2–4). Analogue voltages were measured and processed digitally using a data acquisition system (DAQ).

For continuous-wave measurements, microwaves supplied by a signal generator were passed through a microwave switch and directional coupler before reaching the cryostat microwave feedthrough and then ultimately the sample. The switch was controlled using the DAQ, and the ‘couple’ port of the directional coupler was fed to a detector diode so that microwave power at the directional coupler could be measured directly. These measurements were then used to compensate for frequency-dependent variations in microwave transmission to the cryostat. For pulsed measurements, a microwave mixer, switch and amplifier were placed in series between the signal generator and directional coupler. Measurements using pulsed light and microwaves were timed using a digital delay generator and an arbitrary waveform generator that operated the mixer, switch and AOM.

A permanent magnet in a motorized goniometric mount was used to apply a magnetic field along the *c* axis of the material in experiments where a magnetic field was required (Figs 3 and 4d). Otherwise, no magnet was present.

Sample fabrication. Samples were diced from a 365- μ m-thick wafer of HPSI 4H-SiC purchased from CREE (part number: W4TRD0R-0200), and had dimensions of roughly 2 mm \times 3 mm. Ring-shaped microwave sources 34 μ m (Fig. 1) and 1 mm (Fig. 2) in diameter were patterned on the surfaces of two samples using standard photolithographic techniques. In both cases, a 10/90 nm Ti/Pt metallization was used. A third sample was mounted on top of a 1.8 mm wide by 9 mm long microwave stripline (Figs 3 and 4) made from Rogers Corporation RT/duroid 6002 plated with 1.8 μ m of Au. All microwave devices were connected via wire bonds to the microwave line in the cryostat.

Measurement techniques. A 5 s exposure time was used to collect the data shown in Fig. 1a and b. Spectrally filtered optically detected magnetic resonance data (Fig. 1c–h) were measured by first collecting photoluminescence spectra at each frequency, both with and without microwave power applied to the sample. The latter was subtracted from the former, and a software binning algorithm was used to sum the number of detector counts collected under each ZPL feature (Δ PL). The binning algorithm was designed to ignore counts resulting from phonon sideband emission, and did this by subtracting out any broad background located underneath a sharp ZPL in a spectrum. The exposure time for each spectrum was 19 s, and the final signal (Δ PL/PL) was normalized by the total number of ZPL detector counts measured when no microwaves were applied (PL). Lorentzian fits to the data were used to determine the frequencies of the observed resonances, as discussed in the Supplementary Information.

Time-resolved spin dynamics were measured optically using the femtowatt photoreceiver, which captured all light in the 900–1,700 nm range. Selective manipulation of one defect species over the others was possible because the microwave resonances of each species were at different frequencies, or could be tuned using a magnetic field. In Figs 2–4, Δ PL/PL = $V_{\text{Mod}}/V_{\text{Tot}}$, where V_{Mod} is the magnitude of the \sim 20 Hz modulated component of the output voltage, and V_{Tot} is the total time-averaged output voltage. Because collected light was not spectrally filtered within the 900–1,700 nm range, the time-resolved signals in Figs 2–4 were normalized by the total infrared luminescence collected from the sample. The data points shown in the time-resolved plots are the mean of multiple consecutive quick scans taken at identical conditions. The error bars are the standard deviation of this mean, calculated using the variance of these quick measurements.

Pulse sequences and additional information. All pulse sequences consisted of an initial laser pulse to polarize the defect spins, followed by microwave pulses for spin manipulation, and then a final laser pulse to measure the spins. For T_1 measurements, an additional laser pulse was added between the microwave and final laser pulses in order to keep the optical duty cycle constant during the measurement (see Supplementary Information). For the Rabi, Fig. 2b Ramsey, and T_1 measurements, the relevant pulse sequence was repeated continuously while the microwave driving field was modulated on and off at \sim 20 Hz. For the Fig. 3b Ramsey, Hahn and CPMG measurements, the appropriate sequence was repeated continuously while the final microwave pulse was modulated between positive and negative phases at \sim 20 Hz. The modulated and average photoluminescence components that resulted were measured using the femtowatt photoreceiver via the DAQ.

Additional data, details of the experimental set-up, sample structures, pulse sequences, and fits to data in Figs 3c and 4d are given in the Supplementary Information.

Synthesis of macrocyclic natural products by catalyst-controlled stereoselective ring-closing metathesis

Miao Yu^{1*}, Chenbo Wang^{1*}, Andrew F. Kyle², Pavol Jakubec², Darren J. Dixon², Richard R. Schrock³ & Amir H. Hoveyda¹

Many natural products contain a C=C double bond through which various other derivatives can be prepared; the stereochemical identity of the alkene can be critical to the biological activities of such molecules. Catalytic ring-closing metathesis (RCM) is a widely used method for the synthesis of large unsaturated rings^{1,2}; however, cyclizations often proceed without control of alkene stereochemistry². This shortcoming is particularly costly when the cyclization reaction is performed after a long sequence of other chemical transformations². Here we outline a reliable, practical and general approach for the efficient and highly stereoselective synthesis of macrocyclic alkenes by catalytic RCM; transformations deliver up to 97% of the *Z* isomer owing to control induced by a tungsten-based alkylidene. Utility is demonstrated through the stereoselective preparation of epothilone C (refs 3–5) and nakadomarin A (ref. 6), the previously reported syntheses of which have been marred by late-stage, non-selective RCM^{7–12}. The tungsten alkylidene can be manipulated in air, delivering the products in useful yields with high stereoselectivity. As a result of efficient RCM and re-incorporation of side products into the catalytic cycle with minimal alkene isomerization, desired cyclizations proceed in preference to alternative pathways, even under relatively high substrate concentration.

Catalytic RCM of alkenes is indispensable to the preparation of cyclic structures¹; it is used extensively in the synthesis of biologically active molecules². RCM is broadly used in accessing large rings, despite the lack of a reliably stereoselective variant, the availability of which would substantially enhance the value of this critical class of reactions. The absence of stereochemical control originates from the dependency of the catalytic ring closure on the energetic attributes of the product stereoisomers (rather than being dictated by the catalyst). With small or medium rings, *Z* alkenes are generated exclusively; this is not so with sizeable rings, because, frequently, either the energy difference between the two isomeric alkenes is insufficient for achieving high stereoselectivity by thermodynamic control, or, if one isomer is adequately lower in energy, the catalyst is unable to promote facile equilibration.

The severe shortcoming in the state-of-the-art is illustrated by the two sets of non-selective catalytic RCM, shown in Fig. 1, performed en route to macrocyclic natural products epothilone C^{3–5} and nakadomarin A⁶. Efforts from several laboratories have focused on catalytic RCM for synthesis of the macrocyclic moiety of different members of the epothilone family; popular catalysts, like those derived from alkylidene **1** (ref. 13; Fig. 1) and carbenes **2a–d** (refs 14, 15; Fig. 1), deliver little or no stereoselectivity^{7,8}. Initiatives regarding nakadomarin A (see **5**→**6**, Fig. 1), consisting of four different routes that incorporate a late-stage catalytic macrocyclic ring closure, have met with equally unsatisfactory outcomes^{9–12}.

Catalytic stereoselective RCM of dienes **3** and **5** (Fig. 1) constitutes particularly compelling objectives for several reasons. Epothilone C

(precursor to epothilone A), as well as nakadomarin A, belong to important classes of natural products that exhibit exceptional biological activity^{4–6,16}. Epothilones are potent naturally occurring tubulin polymerization and microtubule stabilizing agents that have been investigated extensively. The geometry of the macrocyclic alkene has been shown to influence their activity¹⁶; the *Z* macrocyclic alkene of epothilone C is needed for the desired stereochemical outcome in the preparation of epothilone A through epoxidation^{7,8}. Nakadomarin A is a potent anti-microbial and anti-cancer agent isolated only in minute quantities^{6,12}. An effective method for laboratory synthesis of such important targets could lead to larger quantities of these molecules or their analogues, which might not be easily accessible by fermentation¹⁷. As synthesis of the large rings in epothilone C or nakadomarin A entails the use of extensively functionalized substrates and occurs late in a multi-step sequence, a non-selective transformation inflicts a costly diminution in efficiency; this difficulty is exacerbated by the fact that the two alkene isomers of epothilone C and nakadomarin A are difficult to separate^{12,18}. Moreover, with structurally complicated dienes such as **3** or **5**, the tactic of carrying out preliminary studies involving simpler structural variants to help establish the feasibility of an RCM strategy is unreliable; substituents and their stereochemical identities have a pivotal role in the efficiency and stereoselectivity of catalytic ring closures and their absence often has a major influence on the cyclization process⁷. In a catalyst-controlled RCM, stereoselectivity would become far less dependent on the attributes of the diene starting material and therefore more predictable.

At the time the present investigations were initiated, efforts to address the above complications had centred on the more common but less efficient detour of altering substrate structure (rather than identification of a catalyst that generates the desired alkene stereoselectively). One relatively established two-step approach consists of W- or Mo-catalysed alkyne RCM followed by Pd-catalysed partial hydrogenation^{19,20}: the first process affords the ring system and the other adjusts the oxidation state. Syntheses of the methyl-substituted alkyne precursors, required to enhance catalyst longevity and avoid oligomerization²¹, necessitate additional manipulations; elevated temperatures (80–140 °C) are required for ring closure, and the presence of Lewis-base alkylamines can lead to a need for high loadings of the metal complex (for example, 50 mol%; ref. 22) or complete catalyst inhibition²³. More recently, macrocyclic RCM of a limited range of substrates involving reactions between an internal vinylsilane and a terminal alkene, followed by protodesilylation, has been disclosed²⁴. Two additional steps are again needed: the requisite vinylsilanes are prepared by Ru-catalysed alkyne hydrosilylation, and the resulting trisubstituted silyl-substituted alkenes are converted to the cyclic *Z* alkene by treatment with a mixture of an ammonium fluoride, a silver fluoride salt and acetic acid. High catalyst loadings (20 mol%) are used in the latter RCM strategy, partly because of the intermediacy of a trisubstituted alkene.

¹Department of Chemistry, Merkert Chemistry Center, Boston College, Chestnut Hill, Massachusetts 02467, USA. ²Department of Chemistry, Chemistry Research Laboratory, University of Oxford, Oxford OX1 3TA, UK. ³Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

*These authors contributed equally to this work.

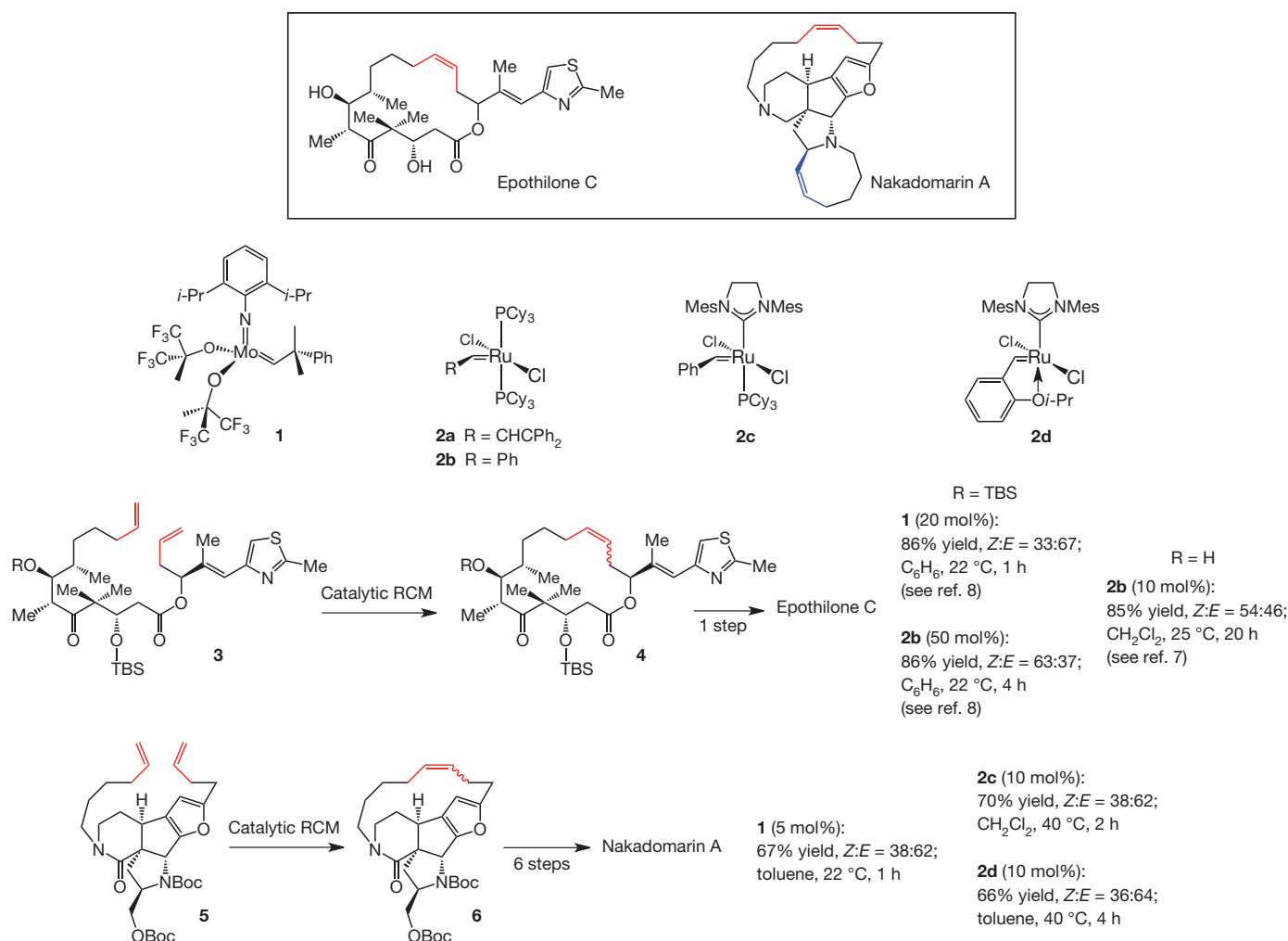


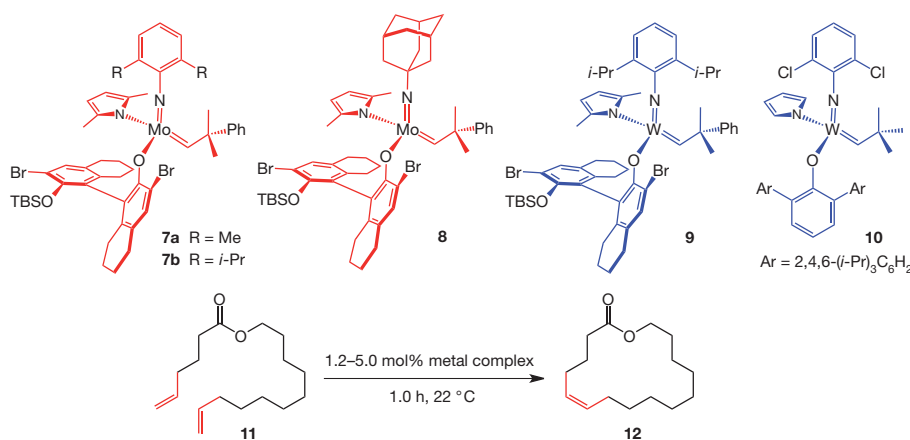
Figure 1 | Two cases in natural product total synthesis where catalytic RCM with commonly used complexes (**1**, **2b–d**) delivers the macrocyclic alkene with minimal stereoselectivity. Top row, target compounds epothilone C and nakadomarin A. Second row, frequently employed complexes **1**, **2b–d**. Third row, catalytic RCM used in synthesis of epothilone C, showing results with different catalysts. Difficulties in stereoselective ring closure are particularly

detrimental because the catalytic RCM takes place late in the synthesis route, inflicting substantial loss in efficiency. For example, diene **3**, used in the total synthesis of anti-cancer agent epothilone C, is prepared by a 16-step sequence. Bottom, catalytic RCM used in synthesis of nakadomarin A. TBS, *t*-butyldimethylsilyl; Boc, *t*-butoxycarbonyl; Mes, 2, 4, 6-trimethylphenyl.

We have introduced several types of intermolecular *Z*-selective olefin metathesis reactions promoted by molybdenum and, less commonly, tungsten alkylidenes that bear a pyrrolide and an alkoxide or an aryloxide ligand. Such stereogenic-at-metal²⁵ catalysts initiate *Z*-selective alkene formation by ring-opening/cross-metathesis²⁶, homocoupling²⁷ or the more complicated cross-metathesis²⁸. Stereochemical models that provide a mechanistic foundation for high *Z* selectivity have been proposed²⁸ and are based on the size differential between the large aryloxide and the smaller imido group (metallacyclobutane substituents oriented towards the latter; see Supplementary Information for details). Successful design of stereoselective macrocyclic RCM reactions, however, requires addressing challenges that are distinct from those pertaining to stereoselective cross-metathesis reactions. When RCM or cross-metathesis involves two unhindered alkenes, stereoisomeric purities can be fragile, because the kinetically generated *Z* alkene can more readily undergo isomerization to the *E* isomer²⁸. With many cyclizations, such as those in Fig. 1, there is no allylic substituent to discourage association of the macrocyclic *Z* alkene with the catalyst to retard the rate of unwanted equilibration; adventitious ring-opening can pose a serious problem. Achieving high stereoselectivity and yield often calls for a catalyst that delivers the subtle and difficult balance that culminates in an efficient and *Z*- or *E*-selective cyclization with little or no ring-opening/ring-closing that can cause isomerization.

Thus, a complicating factor that is critical to RCM but does not apply to cross-metathesis relates to the interplay between ring closure and isomerization by ring-opening. Furthermore, a common strategy in cross-metathesis relates to the use of excess amounts of one cross-partner to favour formation of the desired product (versus homocoupling or isomerization)²⁸; in an RCM, on the other hand, the two reacting alkenes can only be present at the same concentration. Whereas the steric and electronic attributes of one alkene may be rendered distinct in cross-metathesis as the means to minimize homocoupling and enhance the yield of the desired product²⁸, such strategic differentiations are often not possible in RCM (compare Fig. 1). Unlike cross-metathesis, conformational preferences can be critical to the facility of RCM, aiding or resisting the influence of the catalyst. Finally, in cross-metathesis only homocoupling can lead to adventitious substrate consumption, whereas in RCM the same side product continues to deplete the substrate amount through oligomerization.

Examining the ability of different catalysts to promote the RCM of diene **11** to afford the 16-membered-ring lactone **Z-12** (Table 1) was first on our agenda. Preliminary density functional theory (DFT) calculations (Supplementary Information) had revealed that the *E* isomer is 1.2 kcal mol^{−1} lower in energy than the *Z* isomer, suggesting that, at equilibrium, there would exist an approximately 12:88 *Z*:*E* mixture. A previously disclosed attempt involving ruthenium carbene **2a** delivered

Table 1 | Initial study of catalytic RCM of diene 11 to generate 12 stereoselectively

Entry no.	Metal complex	Catalyst loading* (mol%)	Pressure	Conv.† (%); yield‡ (%)	Z:E†
1	1	5.0	Ambient	85; 60	22:78
2	1	5.0	7.0 torr	96; 58	21:79
3	2c	5.0	Ambient	75; 61	21:79
4	7a	5.0	Ambient	56; 45	70:30
5	7a	5.0	7.0 torr	97; 56	77:23
6	7b	5.0	7.0 torr	91; 55	72:28
7	8	3.0	7.0 torr	80; 62	85:15
8	8	1.2	7.0 torr	75; 56	92:8
9	9	5.0	7.0 torr	80; 62	91:9
10	10	5.0	7.0 torr	14; 10	95:5

The reactions were carried out in toluene (5.0 mM) at 22 °C for one hour under an atmosphere of nitrogen gas or under vacuum, as noted; reaction in entry 3 performed in CH₂Cl₂ at 40 °C. See Supplementary Information for details.

* Complexes **1**, **2c**, **9**, **10** were prepared before use; alkylidenes **7a**, **b** and **8** were synthesized *in situ* from the bis-pyrrolide and aryl alcohol, which proceeds in >98% yield for **7a**, **b** but in 60% (±5%) yield in the case of **8** (thus, catalyst loading for the latter complex is 3.0 mol%). See Supplementary Information for details.

† Conversion and Z:E ratios measured by analysis of 400 MHz ¹H NMR spectra of unpurified mixtures; variance of values is estimated to be ±2%.

‡ Yield of isolated product after purification; variance of values is estimated to be <±5%.

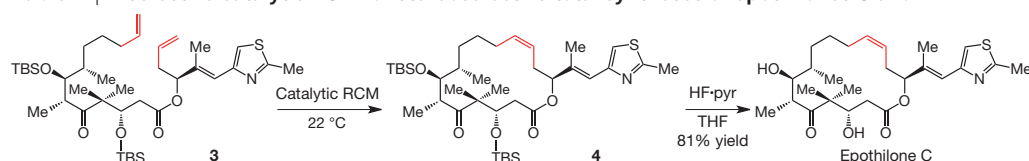
12 with 77% *E* selectivity (that is, 23% *Z*; 4.0 mol%, 22 °C, 30 h)²⁹. As demonstrated in entries 1–3 in Table 1, *E*-**12** is formed preferentially with complexes **1** or **2c**; reduced pressure, a strategy used to minimize isomerization²⁸, does not improve selectivity (compare entries 1 and 2). In contrast, *Z*-**12** is generated with moderate preference when RCM is carried out with monopyrrolides **7a**, **b** (entries 4–6). Adamantylimido **8** furnishes 85% of the *Z* isomer under vacuum (7.0 torr; 62% yield; entry 7); stereoselectivity increases to 92% *Z* with 1.2 mol% catalyst loading (entry 8; versus 3.0 mol% in entry 7) while affording similar efficiency, presumably because isomerization of the cyclic *Z* alkene is reduced when the catalyst is less available. Reaction with tungsten alkylidene **9** leads to equally high yield and stereochemical control (62% and 91% *Z*; entry 9). There is exceptional stereoselectivity with dichloroimido tungsten alkylidene **10** (ref. 27) (95% *Z*; entry 10, Table 1), but the reaction proceeds only to 14% conversion with this less active and sterically more demanding complex; longer reaction times do not result in significantly higher conversion. The preference for generation of the *Z* macrocycle is likely to be due to similar principles that result in stereoselective homocoupling and cross-metathesis reactions (see Supplementary Information for a proposed model).

Next, we turned our attention to the challenge of achieving high *Z* selectivity in RCM reactions that lead to epothilones C and A (compare Fig. 1). We prepared diene **3** along the lines of a formerly devised 16-step sequence⁷. Treatment of **3** with Ru-based **2d** leads to preferential formation of the *E* isomer (66%; entry 1, Table 2). Use of 10 mol% arylimido Mo alkylidene **7a** gives rise to 57% conversion to macrocyclic alkene **4** within three hours, but the *Z* isomer is only generated with 64% selectivity (entry 2). When adamantylimido **8** is used under the same conditions (entry 3), efficiency and stereoselectivity improve (87% conversion in 1.5 h and 85% *Z*), presumably as a result of a more accessible metal centre and larger size differential between the aryloxide and the alkylidene unit (compare the stereochemical model in Supplementary Information). There is only a limited enhancement of conversion and stereoselectivity under reduced pressure (entry

4 versus 3, Table 2). When ring closure is carried out under 1.0 torr of vacuum with tungsten alkylidene **10** (entry 5), which bears a 2,6-dichlorophenylimido and a bulky 2,6-di-[2,4,6-(*i*-Pr)₃]-phenoxy ligand (versus aryloxides in **7**–**9**), the reaction proceeds to near completion in the same amount of time (2.5 h, 97% conversion), allowing the desired macrocycle to be isolated in 85% yield (96% *Z*). As the data in entry 6 of Table 2 illustrate, with the reaction mixture 50 times more concentrated (0.05 M), 3.0 mol% of the same alkylidene can be used to synthesize the desired product (**4**) in 63% yield and 97% *Z* selectivity. The wider gap between percentage conversion and yield values (97% and 63%, respectively) is largely the result of adventitious oligomerization, likely to have been facilitated by the increased substrate concentration. Lactone **4** is converted to epothilone C on silyl ether removal (81% yield; Table 2); epoxidation of epothilone C generates epothilone A^{7,8}.

The higher stereoselectivities furnished by W-based complex **10** are probably because, as stated above, it possesses the desired activity level. The tungsten alkylidene promotes efficient RCM at reasonable catalyst loading without being too active, which would lead it to readily cause *Z*-to-*E* isomerization—even at late stages of the transformation when diene concentration is low. In contrast, the more active Mo-based variants probably initiate a similarly *Z*-selective RCM but also engender subsequent ring-opening/isomerization. It is possibly due to such attenuated reactivity that—contrary to the commonly held perception—tungsten alkylidene **10** proves to be sufficiently stable to be handled in air. An example is shown in entry 7 of Table 2: with 7.5 mol% **10**, weighed in air under up to 80% humidity, and all manipulations performed in a fume hood with standard glassware, macrocyclic alkene **4** is delivered in 82% yield and 94% *Z* selectivity (on the 219 mg scale). It should be mentioned that the faster acting Mo complexes, superior to W-based alkylidenes in effecting intermolecular cross-metathesis reactions²⁸, are more sensitive to air and moisture.

The above findings broach the question as to whether the low stereoselectivity in the synthesis of the 15-membered ring moiety of nakadomarin A (Fig. 1) can be addressed through the use of monopyrrolide

Table 2 | Z-selective catalytic RCM for stereoselective total syntheses of epothilones C and A

Entry no.	Complex; loading (mol%)	Pressure; concentration	Time	Conv.† (%); yield‡ (%)	Z:E†
1	2d ; 5	Ambient; 1.0 mM	16 h	96; ND	34:66
2	7a ; 10	1.0 torr; 1.0 mM	3.0 h	57; ND	64:36
3	8 ; 10	Ambient; 1.0 mM	1.5 h	87; ND	85:15
4	8 ; 10	1.0 torr; 1.0 mM	1.5 h	91; ND	90:10
5	10 ; 10	1.0 torr; 1.0 mM	2.5 h	97; 85	96:4
6	10 ; 3.0	1.0 torr; 0.05 M	3.0 h	97; 63	97:3
7	10 ; 7.5*	0.02 torr; 6.0 mM	4.0 h	96; 82	94:6

The reactions were carried out at 22 °C in purified benzene (under an atmosphere of nitrogen gas) or toluene (vacuum), except for entry 7 (in mesitylene); see Supplementary Information for details. Note that epoxidation of epothilone C gives epothilone A. THF, tetrahydrofuran; ND, not determined.

* Catalyst was weighed in air and reaction performed in a typical fume hood under argon; see Methods Summary for details.

† Conversion and Z:E ratios measured by analysis of 500 MHz ¹H NMR spectra of unpurified mixtures; variance of values is estimated to be $\pm 2\%$.

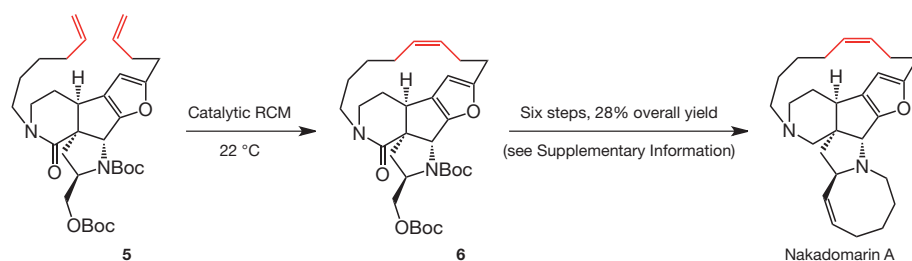
‡ Yield of isolated product after purification; variance of values is estimated to be $\pm 5\%$.

complexes. As indicated by the data in entries 1 and 2 of Table 3, arylimido molybdenum alkylidene **7b** affords only 10% conversion to **6**, a precursor to the natural product. In contrast, the sterically more accessible adamantylimido **8** readily converts **5** to pentacycle **6**, but with only 69:31 Z:E selectivity. Similarly, reaction with tungsten alkylidene **9** is inefficient, probably due to a slow rate of initiation (entry 3, Table 3). The robust tungsten alkylidene **10**, on the other hand, again emerges as the source of a facile and uniquely stereoselective catalyst (entry 4, Table 3): the desired pentacycle **6** is obtained in 90% yield after purification and with 97% Z selectivity (performed with 107 mg of **5**).

It is striking that under conditions (0.1 M; entry 6, Table 3) routinely used to perform a typical chemical transformation (versus high dilution typically required for macrocyclic RCM), reaction proceeds to furnish **6** in 52% yield and 94% Z selectivity. Equally noteworthy is that when cyclization of **5** is carried out at 0.1 M concentration, reduced pressure is not necessary (that is, ethylene is not removed). Otherwise, **6** is obtained in lower yield and selectivity (39% and 90% Z under 7.0 torr, entry 5, Table 3). Since at higher concentration of the diene, homocoupling is rampant, it is likely that the ethylene formed as the by-product raises the availability of the highly reactive methylidene complex²⁸, which converts the homocoupled product to the monomeric RCM substrate, thus increasing the yield of the desired product. The above scenario, and the fact that Z selectivity remains high at 0.1 M concentration (94% Z), implies that the tungsten methylidene

reacts with the acyclic alkene of the homocoupled triene preferably (rather than with the cyclic alkene **6** to promote isomerization). The somewhat lower stereoselectivity observed under vacuum (90:10 versus 94:6 Z:E, entries 5 and 6, Table 3) might be because some macrocyclic product is formed through RCM involving the alkylidene derived from the terminal alkene of the homocoupled by-product. The latter pathway to pentacyclic **6** can be less selective than RCM via diene **5**, arising from reaction of two terminal alkenes. It is consequently as a result of several delicate reactivity preferences that the RCM with complex **10** in a 0.1 M solution can be performed efficiently and selectively.

Total synthesis of nakadomarin A might alternatively be accomplished by a late-stage stereoselective RCM (rather than at an earlier point, as in the pathway in Table 3); such a plan, however, can present additional complications and a non-selective RCM translates to loss of a more valuable advanced intermediate. One route proceeds through the especially demanding RCM (compared to **5**→**6**) of azacene-containing **13** (ref. 12; Fig. 2): the higher ring strain within the pentacyclic diene substrate is not only expected to discourage ring closure, it probably lowers the barrier to undesired rupture of the macrocyclic alkene. Accordingly, past attempts at achieving the transformation of **13** to nakadomarin A, as shown in Fig. 2, bottom, have involved the significantly less reactive ruthenium carbene **2b**¹⁴ (compared to **2c, d**) in order to minimize post-RCM isomerization of the macrocyclic alkene. Use of such a reluctant catalyst, which must be introduced slowly,

Table 3 | Z-selective catalytic RCM for stereoselective synthesis of **6 en route to nakadomarin A**

Entry no.	Complex; loading (mol%)	Pressure; concentration	Time	Conv.* (%); yield† (%)	Z:E*
1	7b ; 5.0	7.0 torr; 5.0 mM	2.0 h	10; ND	ND
2	8 ; 6.0	7.0 torr; 5.0 mM	2.0 h	95; 71	69:31
3	9 ; 5.0	7.0 torr; 5.0 mM	2.0 h	26; ND	ND
4	10 ; 5.0	7.0 torr; 5.0 mM	2.0 h	98; 90	97:3
5	10 ; 5.0	7.0 torr; 0.1 M	0.5 h	>98; 39	90:10
6	10 ; 5.0	Ambient; 0.1 M	2.0 h	95; 52	94:6

The reactions were carried out in purified toluene under an atmosphere of nitrogen gas or under vacuum, as noted. The stereochemical identity of **6** was determined by X-ray crystallography. See Supplementary Information for details. Boc, t-butoxycarbonyl.

* Conversion to macrocyclic alkene **6** and Z:E ratios measured by analysis of 400 MHz ¹H NMR spectra of unpurified mixtures; variance of values is estimated to be $\pm 2\%$.

† Yield of isolated macrocyclic alkene **6** after purification; variance of values is estimated to be $\pm 5\%$.

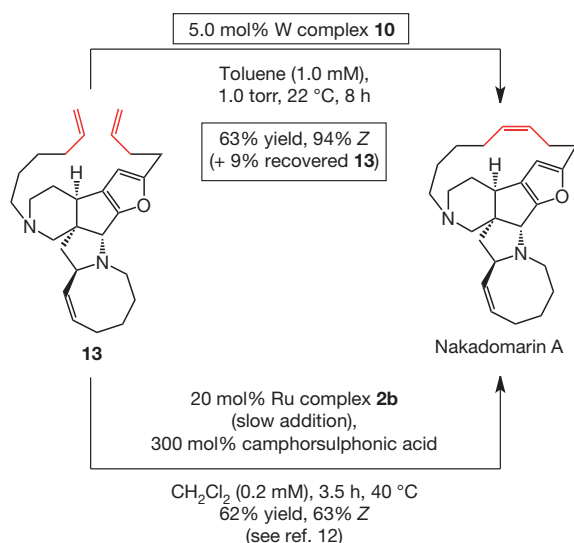


Figure 2 | Total synthesis of nakadomarin A realized through late-stage tungsten-catalysed RCM of pentacyclic **13**, and comparison with results delivered by Ru catalysts. RCM of the strained **13** with tungsten complex **10** affords the natural product in 63% yield (69% based on recovered substrate) and 94% Z-selectivity (top route in figure). This is in contrast to previous attempts, the best of which involves 20 mol% of a Ru carbene added slowly to a highly dilute solution (0.2 mM) to generate only 63:37 Z:E mixture (bottom route in figure).

translates to high loadings and elevated temperatures (20 mol%, 40 °C). Extremely dilute conditions (0.2 mM) are also needed, as it is unlikely that under such conditions any homocoupled by-products that would otherwise be formed can be reverted back to the monomeric dienes or converted directly to the desired macrocycle. Additionally, the presence of substantial quantities (300 mol%) of camphorsulphonic acid, a strong Brønsted acid, is required for achieving 63% Z selectivity (otherwise, slight excess of the E alkene is obtained)¹². In sharp contrast (Fig. 2, top), treatment of **13** with 5.0 mol% **10** at 22 °C affords nakadomarin A in 94% Z selectivity and 63% yield (plus 9% recovered diene). Finally, it should be noted that attempts to effect alkyne RCM of the diyne derivative of **13** (Me-substituted), bearing two Lewis basic tertiary amines, with either Mo- or W-based alkylidynes, leads to <5% conversion even with 30–50 mol% of a metal complex and at 80 °C (up to 24 h); this latter approach must therefore involve the use of the derived bisamide (20–25 mol% catalyst, 80 °C, 16–18 h).

The investigations described above point to stereogenic-at-tungsten alkylidenes as practical and uniquely effective catalysts for Z-selective macrocyclic RCM. We demonstrate that, in planning a multi-step pathway for the preparation of a complex molecule, such complexes can be relied on to deliver the desired outcome at the late stages of an extended route. The impact of stereoselective W-catalysed macrocyclizations reaches beyond the target structures probed in this study, as there are numerous other total syntheses^{22,30} of biologically active molecules that would similarly benefit from the protocols disclosed here.

METHODS SUMMARY

General procedure for catalytic Z-selective macrocyclic RCM. A 250-ml Schlenk flask, fitted with a connecting adaptor attached to an argon-filled manifold, was flame-dried and charged with diene **3** (0.219 g, 0.298 mmol). After azeotropic distillation with dry benzene (three times; freeze-pump), the apparatus was charged with tungsten complex **10** (21.9 mg, 22.4 μmol, weighed in air), evacuated, back-filled with argon and charged with mesitylene (50.0 ml). The mixture was exposed to vacuum (0.02 torr) and allowed to stir for four hours at 22 °C, after which the reaction was quenched by the addition of wet diethyl ether (~1 ml). Purification by silica gel chromatography (hexanes:diethyl ether 20:1) afforded **4** (0.172 g, 0.243 mmol, 82% yield, 94:6 mixture of Z:E isomers, determined by 500 MHz ¹H NMR) as a white foam and 9.3 mg of the recovered starting material (13 μmol, 3.0%).

Received 22 June; accepted 15 September 2011.

- Hoveyda, A. H. & Zhugralin, A. R. The remarkable metal-catalyzed olefin metathesis reaction. *Nature* **450**, 243–251 (2007).
- Gradillas, A. & Perez-Castells, J. Macrocyclization by ring-closing metathesis in the total synthesis of natural products: reaction conditions and limitations. *Angew. Chem. Int. Edn* **45**, 6086–6101 (2006).
- Höfle, G. et al. Epothilone A and B — novel 16-membered macrolides with cytotoxic activity: isolation, crystal structure, and conformation in solution. *Angew. Chem. Int. Edn* **35**, 1567–1569 (1996).
- Kowalski, R. J., Giannakakou, P. & Hamel, E. Activities of the microtubule-stabilizing agents epothilones A and B with purified tubulin and in cells resistant to paclitaxel (taxol). *J. Biol. Chem.* **272**, 2534–2541 (1997).
- Bollag, D. M. et al. Epothilones, a new class of microtubule-stabilizing agents with a taxol-like mechanism of action. *Cancer Res.* **55**, 2325–2333 (1995).
- Kobayashi, J., Watanabe, D., Kawasaki, N. & Tsuda, M. Nakadomarin A, a novel hexacyclic manzamine-related alkaloid from *Amphimedon* sponge. *J. Org. Chem.* **62**, 9236–9239 (1997).
- Nicolaou, K. C. et al. The olefin metathesis approach to epothilone A and its analogues. *J. Am. Chem. Soc.* **119**, 7960–7973 (1997).
- Meng, D. et al. Total synthesis of epothilones A and B. *J. Am. Chem. Soc.* **119**, 10073–10092 (1997).
- Nagata, T., Nakagawa, M. & Nishida, A. The first total synthesis of nakadomarin A. *J. Am. Chem. Soc.* **125**, 7484–7485 (2003).
- Ono, K., Nakagawa, M. & Nishida, A. Asymmetric total synthesis of (–)-nakadomarin A. *Angew. Chem. Int. Edn* **43**, 2020–2023 (2004).
- Young, I. S. & Kerr, M. A. Total synthesis of (+)-nakadomarin A. *J. Am. Chem. Soc.* **129**, 1465–1469 (2007).
- Jakubec, P., Cockfield, D. M. & Dixon, D. J. Total synthesis of (–)-nakadomarin A. *J. Am. Chem. Soc.* **131**, 16632–16633 (2009).
- Schrock, R. R. & Hoveyda, A. H. Molybdenum and tungsten imido alkylidene complexes as efficient olefin metathesis catalysts. *Angew. Chem. Int. Edn* **42**, 4592–4633 (2003).
- Scholl, M., Ding, S., Lee, C. W. & Grubbs, R. H. Synthesis and activity of a new generation of ruthenium-based olefin metathesis catalysts coordinated with 1,3-dimesityl-4,5-dihydroimidazol-2-ylidene ligands. *Org. Lett.* **1**, 953–956 (1999).
- Garber, S. B., Kingsbury, J. S., Gray, B. L. & Hoveyda, A. H. Efficient and recyclable monomeric and dendritic Ru-based metathesis catalysts. *J. Am. Chem. Soc.* **122**, 8168–8179 (2000).
- Altmann, K. H. et al. The total synthesis and biological assessment of *trans*-epothilone A. *Helv. Chim. Acta* **85**, 4086–4110 (2002).
- Starks, C. M., Zhou, Y., Liu, F. & Licari, P. J. Isolation and characterization of new epothilone analogues from recombinant *Myxococcus xanthus* fermentation. *J. Nat. Prod.* **66**, 1313–1317 (2003).
- Schinz, D. et al. Total synthesis of (–)-epothilone A. *Chem. Eur. J.* **5**, 2483–2491 (1999).
- Fürstner, A., Mathes, C. & Lehmann, C. W. Alkyne metathesis: development of a novel molybdenum-based catalyst system and its application to the total synthesis of epothilone A and C. *Chem. Eur. J.* **7**, 5299–5317 (2001).
- Nilson, M. G. & Funk, R. L. Total synthesis of (–)-nakadomarin A. *Org. Lett.* **12**, 4912–4915 (2010).
- Coutelier, O. & Mortreux, A. Terminal alkyne metathesis: a further step towards selectivity. *Adv. Synth. Catal.* **348**, 2038–2042 (2006).
- Smith, B. J. & Sulikowski, G. A. Total synthesis of (±)-haliclonyclammine C. *Angew. Chem. Int. Edn* **49**, 1599–1602 (2010).
- Zhang, W. & Moore, J. S. Alkyne metathesis: catalysts and synthetic applications. *Adv. Synth. Catal.* **349**, 93–120 (2007).
- Wang, Y. et al. Control of olefin geometry in macrocyclic ring-closing metathesis using a removable silyl group. *J. Am. Chem. Soc.* **133**, 9196–9199 (2011).
- Malcolmson, S. J., Meek, S. J., Sattely, E. S., Schrock, R. R. & Hoveyda, A. H. Highly efficient molybdenum-based catalysts for alkene metathesis. *Nature* **456**, 933–937 (2008).
- Ibrahim, I., Yu, M., Schrock, R. R. & Hoveyda, A. H. Highly Z- and enantioselective ring-opening/cross-metathesis reactions catalyzed by stereogenic-at-Mo adamantylimido complexes. *J. Am. Chem. Soc.* **131**, 3844–3845 (2009).
- Jiang, A. J., Zhao, Y., Schrock, R. R. & Hoveyda, A. H. Highly Z-selective metathesis homocoupling of terminal olefins. *J. Am. Chem. Soc.* **131**, 16630–16631 (2009).
- Meek, S. J., O'Brien, R. V., Llaveria, J., Schrock, R. R. & Hoveyda, A. H. Catalytic Z-selective olefin cross-metathesis for natural product synthesis. *Nature* **471**, 461–466 (2011).
- Fürstner, A. & Langemann, K. Macrocycles by ring-closing metathesis. *Synthesis* 792–803 (1997).
- Fürstner, A., Stelzer, F., Rumbo, A. & Krause, H. Total synthesis of the turrianes and evaluation of their DNA-cleaving properties. *Chem. Eur. J.* **8**, 1856–1871 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the United States National Institutes of Health, Institute of General Medical Sciences (grant GM-59426 to A.H.H. and R.R.S.). M. Y. is a John LaMattina graduate fellow, A.F.K. the recipient of an EPSRC-GlaxoSmithKline synthesis studentship, and P.J. an EPSRC postdoctoral fellow. D.J.D. is grateful for an EPSRC leadership fellowship. We thank S. J. Meek, S. J. Malcolmson, R. V. O'Brien, T. J. Mann and E. T. Kiesewetter for discussions; A. R. Zhugralin, S. Torker and D. L. Silverio for DFT calculations; K. Wu for experimental

assistance; and Boston College for providing access to computational facilities. The X-ray facilities at Boston College are supported by the United States National Science Foundation (CHE-0923264).

Author Contributions M.Y. and C.W. were involved in the discovery, design and development of the Z-selective macrocyclic ring-closing metathesis strategies and applications to the natural product syntheses. A.F.K., P.J. and D.J.D. devised routes for and performed enantioselective syntheses of precursors **5** and **13**. R.R.S. and A.H.H. were involved in the discovery and development of the catalysts used in the study;

A.H.H. conceived and directed the investigations and composed the manuscript with revisions provided by M.Y. and C.W.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.H.H. (amir.hoveyda@bc.edu).

Arabian Sea tropical cyclones intensified by emissions of black carbon and other aerosols

Amato T. Evan¹, James P. Kossin^{2,3}, Chul 'Eddy' Chung⁴ & V. Ramanathan⁵

Throughout the year, average sea surface temperatures in the Arabian Sea are warm enough to support the development of tropical cyclones¹, but the atmospheric monsoon circulation and associated strong vertical wind shear limits cyclone development and intensification, only permitting a pre-monsoon and post-monsoon period for cyclogenesis^{1–4}. Thus a recent increase in the intensity of tropical cyclones over the northern Indian Ocean⁵ is thought to be related to the weakening of the climatological vertical wind shear^{3,4}. At the same time, anthropogenic emissions of aerosols have increased sixfold since the 1930s, leading to a weakening of the southwesterly lower-level and easterly upper-level winds that define the monsoonal circulation over the Arabian Sea^{6–9}. In principle, this aerosol-driven circulation modification could affect tropical cyclone intensity over the Arabian Sea, but so far no such linkage has been shown. Here we report an increase in the intensity of pre-monsoon Arabian Sea tropical cyclones during the period 1979–2010, and show that this change in storm strength is a consequence of a simultaneous upward trend in anthropogenic black carbon and sulphate emissions. We use a combination of observational, reanalysis and model data to demonstrate that the anomalous circulation, which is radiatively forced by these anthropogenic aerosols, reduces the basin-wide vertical wind shear, creating an environment more favourable for tropical cyclone intensification. Because most Arabian Sea tropical cyclones make landfall¹, our results suggest an additional impact on human health from regional air pollution.

The South Asian atmospheric brown cloud (ABC) is a 3-km-thick layer of pollution over the northern Indian Ocean (NIO) and Indian subcontinent that results from human emission of aerosols such as black and organic carbon and sulfates². When over water the dominant surface-radiative effect of the ABC is a decrease in solar insolation, which is enhanced significantly by black and organic carbon aerosols^{7–11}. There is a northward gradient of fine-mode aerosol optical depth over the Arabian Sea (Fig. 1a) and therefore a southward gradient in aerosol surface forcing^{7,9,10,12} (that is, negative surface forcing is greater in magnitude to the north). Positive forcing associated with greenhouse gases is offset in the NIO by the sixfold increase in emissions since 1930 (ref. 9), resulting in a southward gradient in sea surface temperature (SST) trends (Fig. 1b) and a decrease in the climatological summertime northward SST gradient⁷. The atmospheric response to this anomalous SST gradient is a weakening of the southerly cross-equatorial surface flow and upper-level tropical easterly jet, which define the Indian monsoon circulation, through anomalously high sea-level pressure to the north and low sea-level pressure to the south^{7,8}.

Once the monsoon onset occurs, very strong vertical wind shear develops across the Arabian Sea, the main factor prohibiting tropical cyclone development during July and August^{1,2}. Thus, despite climatologically warm SST, only two to three tropical cyclones form in the Arabian Sea every calendar year, an average of one storm in the pre-monsoon period (May and June), and one to two storms in the post-monsoon period (August to December)¹.

For a tropical cyclone to intensify, net heat transfer from the ocean to the near-surface air must be sufficient to overcome the frictional drag acting on the cyclone winds in the boundary layer¹³. In the absence of large-scale kinematic forces that would disrupt the physical structure of a cyclone, such as vertical wind shear¹⁴, storm intensity is limited largely by the thermodynamic environment, as described by potential intensity theory¹⁵. As atmospheric kinematic conditions become more favourable for cyclones to develop, those that do form are more likely to achieve maximum intensities close to their potential intensity. Note that the ABC can affect the mean thermodynamic environment by stabilizing regional atmospheric lapse rates and decreasing potential intensity, but this effect is small relative to the regionally high absolute values of potential intensity¹.

One way to evaluate temporal changes in the intensity of tropical cyclones is to examine cyclone lifetime maximum intensity (LMI), defined as the maximum intensity achieved in the lifetime of a storm¹⁶. We calculate LMI for each storm from historical estimates of cyclone wind speeds¹⁷ and then separate the data into the 1979–1996 and 1997–2010 periods, because ten pre-monsoon tropical cyclones formed during

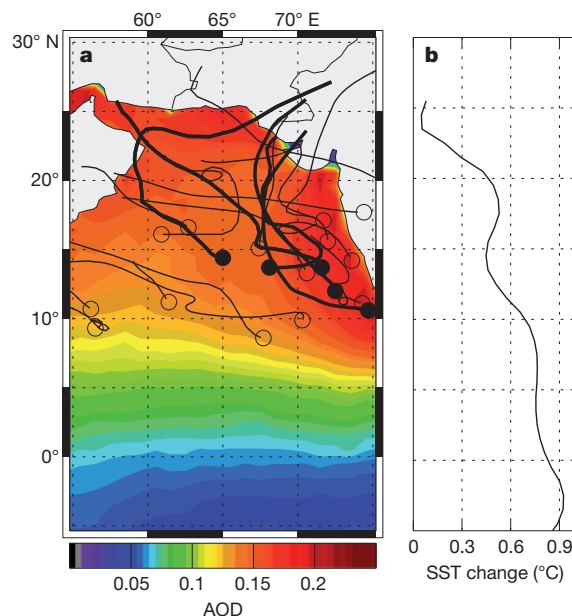


Figure 1 | Tropical cyclone tracks, aerosol optical depth and meridional SST trends in the Arabian Sea. **a**, Genesis points (circles) and tracks (solid lines) of pre-monsoon tropical cyclones during the period 1979–2010. Storms with a lifetime maximum intensity (LMI) of more than 50 m s^{-1} are indicated with a filled circle at the genesis point and thick track lines. Shaded contours represent annual long-term mean fine-mode aerosol optical depth (AOD) from the MODIS Terra and Aqua instruments averaged over 2003–2009. **b**, The 50-year change in observed SST, averaged over 55° – 75° E. The SST change is defined as the average of the monthly linear trend from 1955–2004, multiplied by 50.

¹University of Virginia, Charlottesville, Virginia 22904, USA. ²NOAA's National Climatic Data Center, Asheville, North Carolina 28801, USA. ³NOAA Cooperative Institute for Meteorological Satellite Studies, Madison, Wisconsin 53706, USA. ⁴Gwangju Institute of Science and Technology, Gwangju 500712, Republic of Korea. ⁵Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093, USA.

each epoch. (At the time of writing, data for the 2010 season were not included in the IBTrACS database; track and intensity data for the 2010 Arabian Sea tropical cyclones are therefore from the 2010 Joint Typhoon Warning Center annual tropical cyclone reports (<http://www.usno.navy.mil/JTWC>).) The distribution of pre-monsoon LMI shifts substantially towards greater intensity from the early to the later period, whereas the post-monsoon distribution shifts slightly downwards (Fig. 2a). The difference in the pre-monsoon median (and mean) LMI between the periods is significant at greater than the 97% level, but there is no significant statistical separation for the post-monsoon storms.

Arabian Sea tropical cyclone intensity estimates are based on a Dvorak analysis of satellite imagery¹⁸; given improvements in observational platforms over time it is plausible that the LMI of pre-1998 storms is underestimated, thus causing an apparent increase in the LMI of storms forming after this time. However, if this were true we would expect to see an increase in the LMI of post-monsoon storms, because there is no physical reason why such data artefacts would affect only the pre-monsoon cyclones. Therefore the striking dissimilarity between the changes in distribution of the pre-monsoon and post-monsoon suggests that the change in the pre-monsoon LMI distribution is physical. We repeated this analysis with intensity estimates from an objective analysis of satellite data that is corrected for artefacts related to temporal changes in the observational platform^{16,19}, obtaining nearly identical results (not shown).

We interpret the increasing LMI of pre-monsoon cyclones as reflecting a more favourable kinematic environment for storms to intensify, especially as a persistently high SST helps to maintain a thermodynamic environment that is favourable for storm intensification^{1,3,4}. We speculate that a decrease in the environmental vertical wind shear (defined here as the magnitude of the vector difference between the upper-level and lower-level wind measured along the 200-mbar and 850-mbar pressure levels, respectively) that the storms are experiencing is the catalyst for the increase in pre-monsoon LMI^{1,3,4}.

Using reanalysis data²⁰ we test this hypothesis by calculating the storm-ambient vertical wind shear, defined as the vertical wind shear at the location of the storm averaged over the period during which the tropical cyclone's intensity increases from 17 m s^{-1} to its LMI. In a similar manner to a previously described procedure²¹, we averaged vertical shear values for every cyclone fix 48 h before the arrival of the storm to decrease contamination of the reanalysis fields by the storms themselves. The distribution of pre-monsoon storm-ambient shear shows a pronounced shift towards lower values from the earlier to the later period (Fig. 2b); the median shear value decreases from 11 to 8 m s^{-1} , and the lower quartile for the earlier period is equal to the upper quartile of the later period.

Models²² and observations²³ suggest that the relationship between vertical wind shear and tropical cyclone intensity is not linear; rather, a

tropical cyclone will tend to weaken or strengthen in the presence of vertical shear that is respectively greater than or less than values within the approximate range $8\text{--}11 \text{ m s}^{-1}$. Therefore, as the pre-monsoon shear values become more likely to fall below this range, the storms are more likely to continue intensifying towards their potential intensity and their LMI distribution is expected to shift towards higher values. Although the post-monsoon storm-ambient shear distribution also shifts towards lower values, these values remain mostly within or above this range of thresholds for intensification, and the 1997–2010 post-monsoon shear distribution is not statistically different from the 1979–1996 pre-monsoon shear distribution (Supplementary Table 1). We repeated this analysis using two additional data sets (Supplementary Figs 3–5), obtaining quantitatively similar results (Supplementary Table 1).

We propose that this change in storm-ambient shear is related to a broader downward trend in regional vertical wind shear that results from a simultaneous increase in regional anthropogenic emissions of aerosols. Observational evidence demonstrates that the ABC leads to widespread surface solar dimming of the order of $15\text{--}20 \text{ W m}^{-2}$ over the Arabian sea during the pre-monsoon months¹⁰, and trend analyses of surface solar radiation observations by broad band solar radiometers^{9,24} reveal a nearly 20 W m^{-2} decrease in surface solar radiation from 1950 to 2000, roughly 10% of the climatological solar radiation absorbed at the surface. The dimming is accompanied by a large solar heating of the atmosphere, thus stabilizing the column and further exacerbating the weakening of the circulation resulting from the dimming⁹. We quantify the influence of ABC surface and atmospheric radiative forcing on vertical wind shear by means of numerical experiments designed to identify the response of the regional circulation to the increasing emissions of aerosols over the period 1950–2000 (model experiment details are in Methods).

We estimate the 30-year change in observed pre-monsoon SST by multiplying by 30 the averaged linear trend of observed May and June monthly mean SST²⁵ over the period 1979–2010. To facilitate comparison of SST between the model output and observations we subtract from the observed 30-year SST change the equatorial mean (5° S to 5° N and 55° E to 75° E) 30-year change in SST of 0.3° C (Fig. 3a). We interpret this equatorial change in observed SST as reflecting positive forcing from greenhouse gases. The anomalous 30-year SST change from the model experiments is obtained by differencing the May and June ensemble mean of the perturbation and control experiments. Because the model experiments reflect the SST change from the growth of the optical depth of the ABC over the 50-year period 1950–2000, we assume a linearized response of SST to ABC radiative forcing and scale the model results by three-fifths to estimate the SST change from a 30-year increase in the optical depth of the ABC (Fig. 3b). The observed and modelled southward gradients of the 30-year SST change are

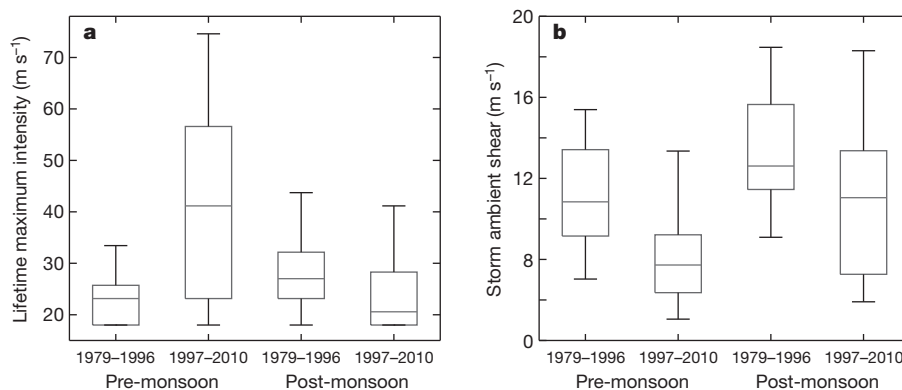


Figure 2 | Distributions of pre-monsoon and post-monsoon LMI and storm-ambient vertical wind shear. Box plots of LMI (a) and storm-ambient vertical wind shear (b) showing the medians (central lines), inner quartile ranges (boxes), and the 25th and 75th centiles minus and plus 1.5 times the

inner quartile range, respectively (whiskers). Shear is calculated from the National Center for Environmental Prediction–Department of Energy Reanalysis²⁰. The significance of the separation of the median and mean values is given in Supplementary Table 1.

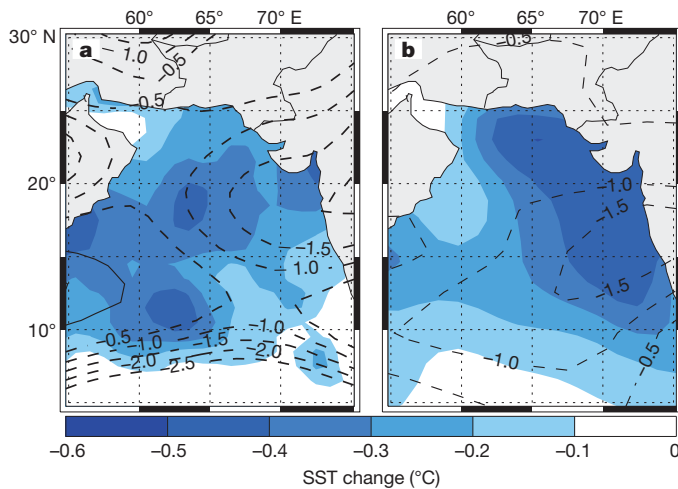


Figure 3 | Thirty-year trends in pre-monsoon SST and vertical wind shear. The 30-year trends in vertical wind shear (contours) based on reanalysis data from 1979–2010 (a) and numerical experiments designed to isolate the effect of the ABC on the regional circulation⁷ (b). Dashed contours indicate negative trends, and solid contours indicate positive trends; the zero contour is not shown. Positive and negative shear contours are in units of 0.5 m s^{-1} . Shading shows the 30-year SST trends over the same period from observations (a), which is relative to the equatorial SST trend, and the aerosol-forced SST change prescribed in the model experiments (b).

similar in magnitude, with SST warming at 5°N being 0.5°C greater than the warming at 15°N in both.

We estimate the historical 30-year change in pre-monsoon vertical wind shear by multiplying by 30 the averaged May and June linear trend of vertical shear from reanalysis²⁰ (Fig. 3a). We assume a linear response of the atmosphere to growth of the ABC and estimate the ABC-forced 30-year change in pre-monsoon vertical wind shear as the difference in ensemble mean May and June vertical shear, scaled by three-fifths (Fig. 3b). The reanalysis and modelled 30-year change in pre-monsoon vertical wind shear are in good agreement north of 10°N , where the ABC is present and where most pre-monsoon Arabian Sea tropical cyclones form and track (Fig. 1a). Each data point shows a 1.5 m s^{-1} decrease off the western coast of the Indian subcontinent, and northward and westward from this location the change in shear in each is 1.0 m s^{-1} . This agreement is compelling evidence that the basin-wide decrease in vertical wind shear is causally related to the increasing intensity of the ABC. We compared the model output with trends in vertical shear from two additional reanalysis data sets, obtaining nearly identical results (Supplementary Fig. 6).

Noting that the distribution of post-monsoon storm-ambient shear and LMI during the 1997–2010 period is similar to those pre-monsoon distributions for the 1979–1996 time span (Fig. 2 and Supplementary Table 1), it can be argued that a further decrease in basin-wide shear associated with projected growth in emissions⁹ will further shift the inner quartile range of the post-monsoon storm-ambient shear distribution towards that of the pre-monsoon shear during the latter period. In such a case it is plausible that very intense tropical cyclones, which have so far been limited to the pre-monsoon period, could begin to emerge in the post-monsoon season as well.

Table 1 | Characteristics of the five most powerful Arabian Sea tropical cyclones, 1980–2010

Cyclone name	Year	LMI (m s^{-1})	Landfall location	Lives lost	Number of people affected	Damage ($\times 10^6$ 2011 US\$)
O3A	1998	54	Gujarat, India	2,871	2,871	643
O2A	1999	57	Karachi, Pakistan	682	576,636	42
O1A	2001	57	Gujarat, India	None	None	None
Gonu	2007	75	Muscat, Oman; Ras Al Kuh, Iran	88	180,009	4,203
Phet	2010	64	Muscat, Oman; Karachi, Pakistan	39	4,000	1,845

Data are shown for the five most powerful Arabian Sea cyclones between 1980 and 2010, all of which formed during the pre-monsoon period. Landfall locations, fatalities, numbers affected and damage data are from the Joint Typhoon Warning Center's annual tropical cyclone reports and the EM-DAT International Disaster Database.

Given the relatively small size of the Arabian basin, more than half of all cyclones that form here make landfall, and even weak landfalling Arabian Sea cyclones can cause considerable destruction and loss of life¹ (see also the Joint Typhoon Warning Center annual tropical cyclone reports, and the EM-DAT International Disaster Database (<http://www.emdat.be/disaster-list>)). All of the most powerful cyclones in the Arabian Sea that occurred during the past 30 years formed during the pre-monsoon period and made landfall in India, Pakistan, Oman or Iran, causing considerable loss of life and substantial damage (Table 1). In addition to the multitude of known human health impacts associated with aerosols that comprise the ABC²⁶, we suggest that the increasing intensity of landfalling tropical cyclones is a consequence of regional emissions of pollution aerosols. Because tropospheric aerosols have a short residence time, decreasing emissions should have a nearly immediate effect on the propensity of pre-monsoon tropical cyclones to reach their maximum potential intensity. However, we caution that this study is conducted from a limited sample of storms; continued research that tests the relationship between aerosols and tropical cyclones in the NIO is warranted.

METHODS SUMMARY

The numerical experiments used here are identical to those described previously⁷, and we refer the reader to that paper and the online Methods for a detailed discussion of the model setup. These experiments were performed using an atmospheric general circulation model²⁷. In the control and perturbation experiments an SST annual cycle is prescribed. In the control experiment SST is prescribed as observed mean SST²⁵ changes from 1950 to 2000. In the perturbation experiment SST in the NIO is estimated to be the equivalent 50-year change in SST assuming that regional emissions had been stabilized at 1950s levels, so that SST is everywhere the same as in the control experiment, except in the NIO, where the local 50-year SST change is forced to be equal to the SST change in the equatorial Indian Ocean⁹. SST in the NIO is warmer in the perturbation run than it is in the control run, and the difference in the zonally averaged (Indian Ocean) meridional SST gradient in the experiments is nearly identical to that from Fig. 1b, except that in the model the difference is zero at the Equator and negative to the north. The perturbation experiment also includes the seasonally varying effect of heating in the atmosphere by the ABC²⁸. We interpret the difference between the 50-member ensemble mean of the control and perturbation experiments as the influence of the 50-year growth in regional emissions on the atmospheric circulation.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 May; accepted 7 September 2011.

1. Evan, A. T. & Camargo, S. J. A climatology of Arabian Sea cyclonic storms. *J. Clim.* **24**, 140–158 (2011).
2. Gray, W. M. Global view of the origin of tropical disturbances and storms. *Mon. Weath. Rev.* **96**, 669–700 (1968).
3. Krishna, K. M. Intensifying tropical cyclones over the North Indian Ocean during summer monsoon—global warming. *Global Planet. Change* **65**, 1–2, 12–16 (2009).
4. Rao, V. B., Ferreira, C. C., Franchito, S. H. & Ramakrishna, S. S. V. S. In a changing climate weakening tropical easterly jet induces more violent tropical storms over the north Indian Ocean. *Geophys. Res. Lett.* **35**, L15710 (2008).
5. Singh, O. P., Kahn, A. & Rahman, S. Has the frequency of intense tropical cyclones increased in the North Indian Ocean? *Curr. Sci.* **80**, 575–580 (2001).
6. Dash, S. K., Kulkarni, M. A., Mohanty, U. C. & Prasad, K. Changes in the characteristics of rain events in India. *J. Geophys. Res.* **114**, D10109 (2009).
7. Chung, C. E. & Ramanathan, V. Weakening of North Indian SST gradients and the monsoon rainfall in India and the Sahel. *J. Clim.* **19**, 2036–2045 (2006).

8. Meehl, G., Arblaster, J. & Collins, W. Effects of black carbon aerosols on the Indian monsoon. *J. Clim.* **21**, 2869–2882 (2008).
9. Ramanathan, V. *et al.* Atmospheric brown clouds: impacts on South Asian climate and hydrological cycle. *Proc. Natl Acad. Sci. USA* **102**, 5326–5333 (2005).
10. Ramanathan, V. *et al.* The Indian Ocean experiment: an integrated assessment of the climate forcing and effects of the great Indo-Asian haze. *J. Geophys. Res.* **106**, 28371–28398 (2001).
11. Wang, C., Kim, D., Ekman, A. M. L., Barth, M. C. & Rasch, P. J. Impact of anthropogenic aerosols on Indian summer monsoon. *Geophys. Res. Lett.* **36**, L21704 (2009).
12. Chung, C. E., Ramanathan, V., Kim, D. & Podgorny, I. Global anthropogenic aerosol direct forcing derived from satellite and ground-based observations. *J. Geophys. Res.* **110**, D24207 (2005).
13. Emanuel, K. Tropical cyclones. *Annu. Rev. Earth Planet. Sci.* **31**, 75–104 (2003).
14. DeMaria, M. & Kaplan, J. An updated statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic and eastern North Pacific basins. *Weather Forecast.* **14**, 326–337 (1999).
15. Emanuel, K. A. The maximum intensity of hurricanes. *J. Atmos. Sci.* **45**, 1143–1155 (1988).
16. Elsner, J. B., Kossin, J. P. & Jagger, T. H. The increasing intensity of the strongest tropical cyclones. *Nature* **455**, 92–95 (2008).
17. Knapp, K. P., Kruk, M. C., Levinson, D. H., Diamond, H. J. & Neumann, C. J. The International Best Track Archive for Climate Stewardship (IBTrACS): unifying tropical cyclone data. *Bull. Am. Meteorol. Soc.* **91**, 363–376 (2010).
18. Dvorak, V. F. *Tropical Cyclone Intensity Analysis using Satellite Data*. Technical Report no. 11 (NOAA, 1984).
19. Kossin, J. P., Knapp, K. R., Vimont, D. J., Murnane, R. J. & Harper, B. A. A globally consistent reanalysis of hurricane variability and trends. *Geophys. Res. Lett.* **34**, L04815 (2007).
20. Kanamitsu, M. *et al.* NCEP-DOE AMIP-II reanalysis (R-2). *Bull. Am. Meteorol. Soc.* **83**, 1631–1643 (2002).
21. Kossin, J. P. & Camargo, S. J. Hurricane track variability and secular potential intensity trends. *Clim. Change* **97**, 329–337 (2009).
22. Frank, W. M. & Ritchie, E. A. Effects of vertical wind shear on the intensity and structure of numerically simulated hurricanes. *Mon. Weath. Rev.* **129**, 2249–2269 (2001).
23. Elsberry, R. L. & Jeffries, R. A. Vertical wind shear influences on tropical cyclone formation and intensification during TCM-92 and TCM-93. *Mon. Weath. Rev.* **124**, 1374–1387 (1996).
24. Kumari, P. B., Londhe, A. L., Daniel, S. & Jadhav, D. B. Observational evidence of solar dimming: offsetting surface warming over India. *Geophys. Res. Lett.* **34**, L21810 (2007).
25. Rayner, N. A. *et al.* Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**, 4407 (2003).
26. UNEP. *Integrated Assessment of Black Carbon and Tropospheric Ozone - Summary for Decision Makers*. UNEP/GC/26/INF/20 (United Nations Environment Program and World Meteorological Organization, 2011).
27. Kiehl, J. T. *et al.* The National Center for Atmospheric Research Community Climate Model: CCM3. *J. Clim.* **11**, 1131–1149 (1998).
28. Chung, C. E. & Ramanathan, V. South Asian haze forcing: remote impacts with implications to ENSO and AO. *J. Clim.* **16**, 1791–1806 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Partial funding for this work was provided by National Oceanic & Atmospheric Administration (NOAA)/Climate Program Office (NA100AR4310136), Korea's Research Agency for Climate Science (RACS 2010-2603) and the National Science Foundation (ATM-0721142). Data from National Centers for Environmental Prediction reanalyses 1 and 2 were provided by the NOAA/Office of Oceanic and Atmospheric Research/Earth System Research Laboratory Physical Sciences Division, Boulder, Colorado, USA, from their website (<http://www.esrl.noaa.gov/psd/>).

Author Contributions A.E. and V.R. conceived the project. A.E. and J.K. designed the study. A.E., J.K. and E.C. provided model and observational data. All authors participated in data interpretation and co-wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.E. (ate9c@virginia.edu).

METHODS

The numerical experiments used in this study to identify the effect of increasing emissions on vertical wind shear are identical to those described previously⁷, and we refer the reader to that paper for a comprehensive discussion of the experimental setup and model results. These numerical experiments were performed with the National Center for Atmospheric Research Community Climate Model Version 3 (NCAR/CCM3)²⁷ at a T42/L18 resolution with prescribed SST. For each of the control and perturbation experiments there were ten ensemble member simulations, all ten with different initial conditions. For each of the ten ensemble simulations the model was spun up to equilibrium for one year, and then run for an additional five years. The ten 5-year simulations were averaged, yielding a 50-year sample for each of the ensemble means considered here.

In the control run and perturbation run, an SST annual cycle is prescribed. In the control experiment, SST is prescribed to globally reflect observed long-term SST²⁵ changes from 1950 to 2000. This is achieved by calculating the monthly linear SST trend over the period 1951–2002, multiplying the trend by 50 (°C per 50 years), and adding this 50-year SST change to the observed monthly SST climatology (calculated over the period 1951–2002). Therefore, in the control experiment, SST in the NIO reflects a 50-year increase in emissions and negative surface forcing by the

ABC. In the perturbation experiment, SST in the NIO is estimated to be the equivalent 50-year change in SST assuming that regional emissions had been stabilized at 1950s levels, such that SST is everywhere the same as in the control experiment, except in the NIO, where the local 50-year SST change is forced to be equal to the SST change in the equatorial Indian Ocean, consistent with previous findings⁹.

SST in the NIO is warmer in the perturbation run than it is in the control run, and the difference in the zonally averaged (Indian Ocean) meridional SST gradient in the experiments is nearly identical to that from Fig. 1b, except that in the model the difference is zero at the Equator and negative to the north. This difference in SST is equivalent to the radiatively forced change in SST from the increase in emissions over the same 50-year period, as demonstrated previously¹⁰. The perturbation experiment also includes the effect of heating in the atmosphere by the absorption of short-wave and long-wave radiation by the ABC; consistent with ref. 28, a decrease in surface solar insolation was prescribed, as was atmospheric heating from the surface boundary layer to a height of roughly 3 km. The seasonality of the surface forcing and imposed heating is consistent with that of the ABC⁹. We interpret the difference between the ensemble mean of the control and perturbation experiments as being the influence of the 50-year growth in regional emissions on the atmospheric circulation.

Highly specialized mammalian skulls from the Late Cretaceous of South America

Guillermo W. Rougier¹, Sebastián Apesteguía^{2,3} & Leandro C. Gaetano^{3,4}

Dryolestoids are an extinct mammalian group belonging to the lineage leading to modern marsupials and placentals^{1,2}. Dryolestoids are known by teeth and jaws from the Jurassic period of North America and Europe^{2,3}, but they thrived in South America up to the end of the Mesozoic era and survived to the beginnings of the Cenozoic^{2,4–7}. Isolated teeth and jaws from the latest Cretaceous of South America provide mounting evidence that, at least in western Gondwana, dryolestoids developed into strongly endemic groups by the Late Cretaceous^{4–9}. However, the lack of pre-Late Cretaceous dryolestoid remains made study of their origin and early diversification intractable. Here we describe the first mammalian remains from the early Late Cretaceous of South America, including two partial skulls and jaws of a derived dryolestoid showing dental and cranial features unknown among any other group of Mesozoic mammals, such as single-rooted molars preceded by double-rooted premolars, combined with a very long muzzle, exceedingly long canines and evidence of highly specialized masticatory musculature. On one hand, the new mammal shares derived features of dryolestoids^{1–3} with forms from the Jurassic of Laurasia, whereas on the other hand, it is very specialized and highlights the endemic, diverse dryolestoid fauna from the Cretaceous of South America. Our specimens include only the second mammalian skull known for the Cretaceous of Gondwana, bridging a previous 60-million-year gap in the fossil record, and document the whole cranial morphology of a dryolestoid, revealing an unsuspected morphological and ecological diversity for non-triassic mammals.

Class Mammalia Linnaeus, 1758
Clade Cladotheria McKenna, 1975
Superorder Dryolestoidea Butler, 1939
Order Meridiolestida nov.
Cronopio dentiactutus gen. et sp. nov.

Etymology. *Cronopio* after the fictional, bizarre creatures central to many stories by Julio Cortázar; *dentiactutus*, from Latin, meaning sharp, acute teeth.

Holotype. MPCA 454 (Museo Provincial Carlos Ameghino, Cipolletti, Río Negro Province, Argentina), an incomplete skull preserving a full upper dentition, but missing the braincase, part of the rostral roof and the lower jaws (Figs 1–4).

Locality and horizon. Candeleros Formation (Cenomanian), Neuquén Group¹⁰. The specimens come from the La Buitrera locality, Río Negro Province, Argentina, that has yielded abundant skeletons of small vertebrates^{11,12} and dinosaurs^{13,14}. See Supplementary Information for further data on age, systematics and morphology.

Diagnosis and description. Medium-sized dryolestoid, with an extremely elongated rostrum, extended edentulous portion of the premaxilla, and orbits placed dorsally because of a tall zygoma with a deeply excavated masseteric fossa. Dental formula I2/?, C1/1, P4/3+, M3/3. *Cronopio* is diagnosed by a very long canine, P¹ separated from

P² by an extensive diastema, and highly mesiodistally compressed molariforms. Upper and lower premolars are supported by two roots whereas the molars rest on a single labiolingually broad root. The lower jaw (Figs 1–2 and 4) possesses a masseteric process that juts out laterally forming a broad platform whereas the angular process is small and medially inflected. *Cronopio* differs from other South American dryolestoids like *Leonardus cuspidatus*^{4,8} in having a contact between the stylocone and paracrista, with a stylocone not as large, the presence of a distinct small parastylar hook and at least one less molar. *Cronopio* and other South American dryolestoids^{4,7,15} share the presence of a posterior premolar (P³ in *Cronopio*) with a small anterior cusp supported by a circular root and a distinct mesiodistal basin supported by a transverse one; this morphology has been interpreted previously either as distinct taxa⁴ or as deciduous teeth^{2,3}. *Cronopio* and other South American dryolestoids share with all other dryolestoids a much taller labial crown height than lingual in the lower molars (the reverse in the uppers), a small talonid, a relatively transverse metacristid, and mesiodistally compressed trigonids^{1–3} (Figs 1–4).

Cronopio's canine is extremely long and moderately compressed buccolingually. The dryolestoid *Drescheratherium* from the Jurassic of Portugal¹⁶ also shows very long canines but not to the extent of *Cronopio*. A distinctive feature of *Cronopio* and several South American dryolestoids is the presence of complex penultimate and fully molarized ultimate premolars^{7–15}. The penultimate premolar coincides with a sharp constriction of the rostrum, leaving a much lower anterior accessory cusp isolated from the moderately basined posterior half of the tooth. This unusual morphology, although with a better developed basin, served as the basis for the recognition of several Late Cretaceous taxa such as *Barberenia* and *Quirogatherium*^{5,6}, which in light of the new evidence probably represent just the penultimate premolar position of other taxa^{4,5,17} described based on molars. Similarly shaped teeth are known for the mesungulatids *Coloniatherium*⁸, *Mesungulatum*^{4–6,17} and *Peligrotherium*^{7,15}, although only for *Peligrotherium* is the position of the tooth P² known unambiguously, indicating that this is a conspicuous feature of many South American dryolestoids. Previous studies^{5,17,18} have recognized two species of triconodont mammals in the Late Cretaceous Los Alamitos Formation; the morphology of the teeth referred to these species is similar to that of the anterior upper and lower premolars of *Cronopio* and *Peligrotherium*. We suggest that these putative triconodonts represent anterior premolar positions of dryolestoids (see Supplementary Information). Upper and lower molars of *Cronopio* (Fig. 4) are highly compressed mesiodistally and are single-rooted. The upper crowns are dominated by a paracone and a stylocone connected to the paracrista. Both, premolars and molars are arranged forming a gentle arch that curves posteriorly along the maxillary border. The simultaneous presence of single-rooted molars and double-rooted premolars is a unique feature distinguishing *Cronopio* from any other Mesozoic mammal.

Cronopio's skull shows a combination of primitive mammalian features and highly specialized traits. *Cronopio* shares with basal

¹Department of Anatomical Sciences and Neurobiology, University of Louisville, Louisville, Kentucky 40202, USA. ²CEBBAD - Fundación de Historia Natural 'Félix de Azara', Universidad Maimónides, Hidalgo 775, Buenos Aires (1405), Argentina. ³CONICET, Consejo Nacional de Investigaciones Científicas y Técnicas, Av. Rivadavia 1917, C1033AAJ, Buenos Aires, Argentina. ⁴Departamento de Cs. Geológicas, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón II, C1428EHA, Buenos Aires, Argentina.



Figure 1 | *Cronopio denticutatus* MPCA 454 type skull and MPCA 453 right lower jaw. **a–c**, MPCA 454 skull in ventral (**a**) and dorsal views (**b**); labial view of the left upper dentition (**c**). **d, e**, *Cronopio denticutatus* MPCA 453 right lower jaw in labial and lingual views (**d**) and detail of the left petrosal and basicranium (**e**).

mammaliaforms features such as the presence of a septomaxilla, an anterior lamina of the petrosal, a lateral flange, and all the elements of cranial circulation thought to be primitive for therians, including a fully developed stapodial system¹⁹. The snout of *Cronopio* is unusually long and narrow with a relatively large orbit, a small temporal area and a robust zygoma deeply excavated by a masseteric fossa. The craniodental morphology, in particular the low coronoid process and the hyperdeveloped masseteric fossa and process, indicates that *Cronopio* emphasized rotation of the jaws during mastication at the expense of a powerful bite²⁰. With the exception of a fully definitive mammalian middle ear²¹, *Cronopio* shows that the major changes in skull morphology characterizing the hypothetical common ancestor of marsupials and placentals²² were not yet established in dryolestoids and that they retained not only a primitive non-tribosphenic dentition, but also a primitive braincase and ear region.

Our phylogenetic analysis (Fig. 5) including a wide array of early mammals identifies *Cronopio* as a member of a cluster of South American dryolestoids (see Supplementary Information). *Leonardus* and other Cretaceous South American dryolestoids are successive sister groups reflecting the establishment of a Cretaceous endemic biota^{4,5}. Although in the Cretaceous of Laurasia marsupial and placental (therians) relatives were major components of the fauna, they are yet to be unequivocally identified in the admittedly sparse localities of South America^{2,23}. Depending on how some controversial fossils from

Australia and Africa are interpreted^{24,25}, tribosphenic mammals would also be originally absent from those continents during the Jurassic and earliest Cretaceous, despite Africa–Europe contacts during the Cretaceous^{2,26}. However, an Early Cretaceous connection between Africa and Europe^{2,26} might have allowed some basal tribosphenic forms to invade northern Africa²⁶. Dryolestoids are present in all bona fide Late Cretaceous localities where mammals are known. This radiation of non-tribosphenic forms was well underway at least by the Cenomanian as evidenced by the highly derived *Cronopio*, its lineage of sharp-toothed insectivores persisting to the latest Cretaceous, as indicated by *Leonardus*. Other South American dryolestoids like the mesungulatids developed instead massive bunodont cusps and cingula, adapted to a sophisticated omnivorous/herbivorous diet, and survived into the Palaeocene, forming part of the earliest communities dominated by South American ungulates^{15,17}. The poorly known South American dryolestoid *Groebertherium*, however, clusters with northern continent dryolestoids and underscores the Pangeic and Jurassic origin of the South American dryolestoid biota, a conclusion also borne out by the extensive ghost lineages implied by the phylogeny presented here (Fig. 5).

Tribosphenic molars are held to be a key evolutionary achievement of marsupials, placentals and their closest relatives, that unlocked a wide range of adaptations enabling the colonization of a variety of ecological niches². *Cronopio* as well as the South American Late

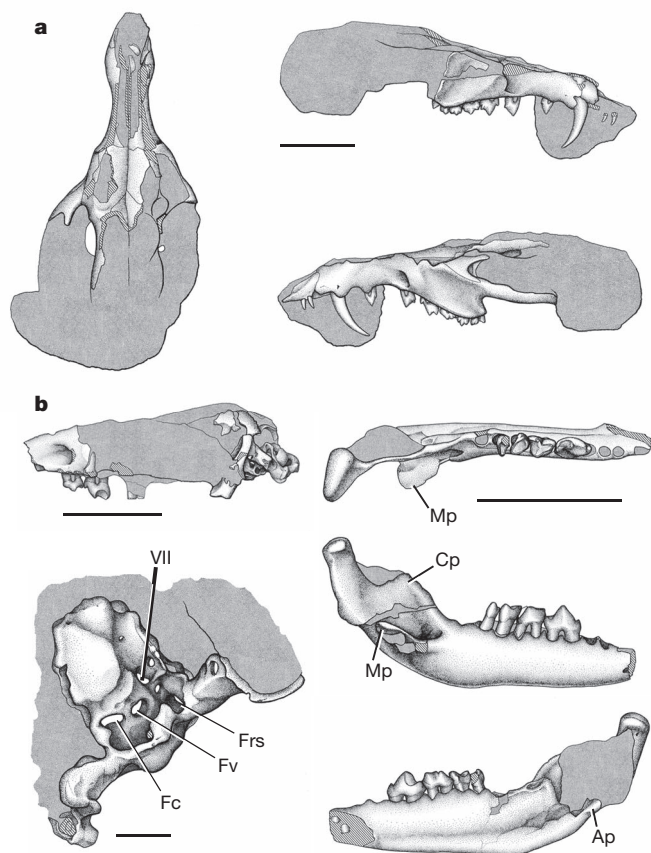


Figure 2 | *Cronopio dentiacutus* MPCA 454 type skull and MPCA 453 skull and right lower jaw. **a**, MPCA 454 skull in dorsal (left) right lateral (top right) and left lateral views (bottom right). **b**, *Cronopio dentiacutus* MPCA 453 left lateral view of the skull (top left), detail of the petrosal and basicranium (bottom left) and occlusal, labial and lingual views of the right lower jaw (right). Ap, angular process; Cp, coronoid process; Fc, fenestra cochleae; Fv, fenestra vestibuli; Frs, foramen for the ramus superior of the stapedial artery; Mp, masseteric process; VII, facial foramen, for hyomandibular branch of cranial nerve VII. All scale bars are 5 mm, except for bottom-left panel of **b**, which is 2 mm.

Cretaceous dryolestoids show that we have underestimated both the biological diversity and morphological disparity of the pre-tribosphenic mammals, which present highly derived mammalian morphotypes previously unknown.

The equally specialized Cretaceous and/or Palaeocene ferugliotheriid and gondwanatherian^{27,28} mammals indicate that all Late Cretaceous mammalian groups of South America are endemic to some degree, but how representative dryolestoids and gondwanatheres are of the late Mesozoic faunas of the non-South-American portions of Gondwana is yet unclear. Isolated dental elements from the Late Cretaceous of Africa²⁹ and North America³⁰ lack *Cronopio*'s autapomorphies, but resemble dryolestoids from the terminal Late Cretaceous of Argentina. At present, the evaluation of those highly fragmentary materials is problematic.

Cronopio is an integral part of the evolutionary history leading to the South American mammalian communities of the Late Cretaceous; however, it shares no close ties with prototribosphenidan, australosphenidan or triconodont mammals from the older Middle Jurassic and Early Cretaceous localities of South America. This faunal difference points to a minimum of one major faunal change in South America between the Middle Jurassic and the early Late Cretaceous. *Cronopio* and other mammals yet to be described from La Buitrera locality are, because of their good preservation and relative completeness, our best hope to address the origin and early evolution of the Mesozoic mammalian faunas of South America.

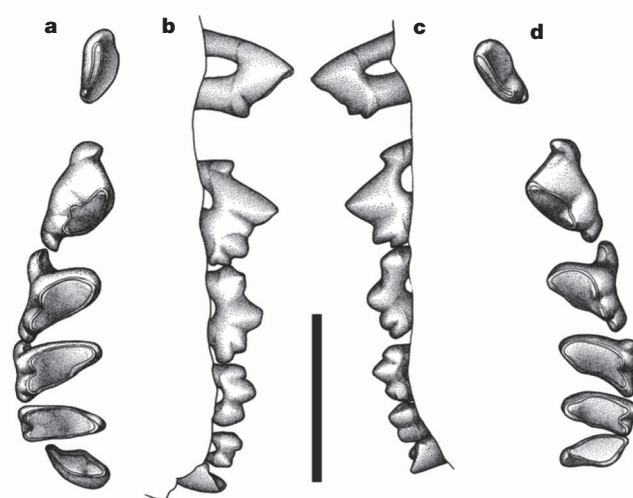


Figure 3 | *Cronopio dentiacutus* MPCA 454 holotype (upper teeth). **a–d**, Upper teeth in occlusal (**a**, **d**) and labial view (**b**, **c**). Scale bar, 3 mm.

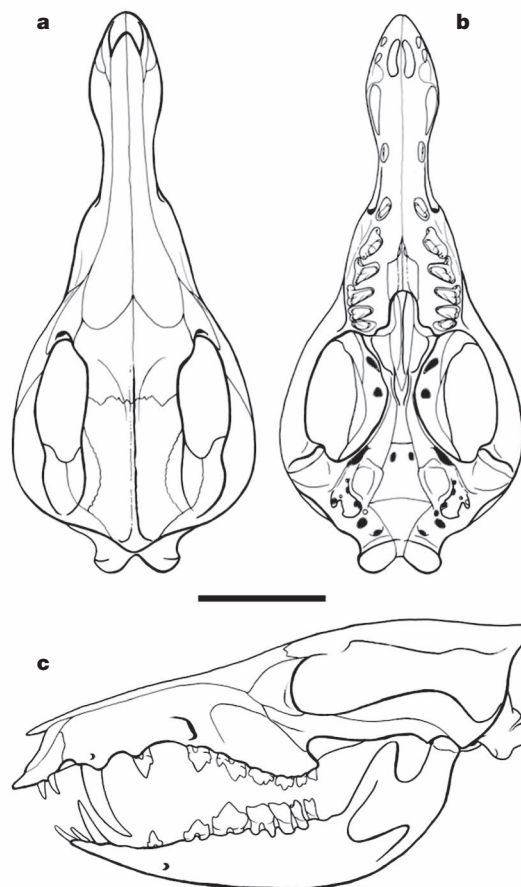


Figure 4 | Reconstruction of the skull and jaws of *Cronopio dentiacutus* based on MPCA 453 and 454. The only areas not preserved in the specimens described here are the lower canines and lower jaw incisors, the septomaxilla, the sagittal crest and part of the cranial vault in the skull. Scale bar, 5 mm. Estimated skull length of MPCA 454 is 27 mm.

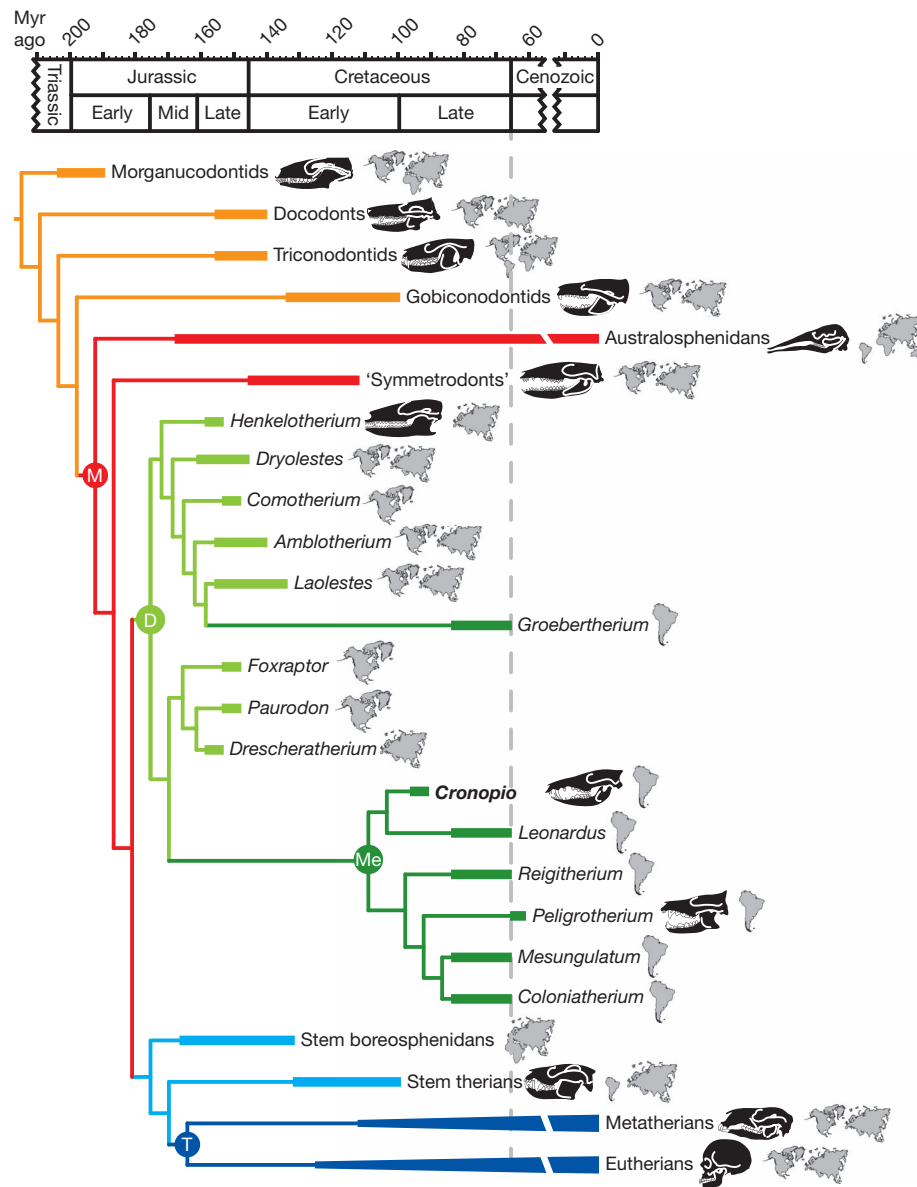


Figure 5 | Simplified phylogenetic tree of the major taxa included in the analysis. The topology reflects a strict consensus tree of 8 equally most parsimonious trees (length 1,114) performed on a matrix including 304 characters and 57 taxa. The dryolestoids (green) show a basal dichotomy with most of the South America forms (dark green) forming a monophyletic group that includes as terminal taxa highly specialized relatively large-sized forms from the Late Cretaceous and Palaeocene. Unlike most other dryolestoids, these derived South American mammals were omnivorous/herbivorous with

complex tooth-on-tooth occlusion and thick enamel, and group together under Mesungulatoidea (see Supplementary Information). *Groebertherium* from the latest Cretaceous of Argentina appears as a derived member of a northern group of dryolestoids, indicating a probable Jurassic origin from groups of Pangeic distribution for at least some Late Cretaceous South American dryolestoids. Letters at the nodes indicate high-level clades: D, Dryolestoidae; M, Mammalia; Me, Meridiolestida; T, Theria. See Supplementary Information for character list, data matrix and analysis protocol. Myr, million years.

METHODS SUMMARY

Commonly available phylogenetic software (Winclada-Asado, NONA, TNT) was used during the study; see Supplementary Information for search parameters and data sources.

Received 15 August; accepted 20 September 2011.

- Prothero, D. R. New Jurassic mammals from Como Bluff, Wyoming, and the interrelationships of non-tribosphenic Theria. *Bull. Am. Mus. Nat. Hist.* **167**, 281–317 (1981).
- Kielan-Jaworowska, Z., Cifelli, R. L. & Luo, Z. X. *Mammals from the Age of Dinosaurs. Origins, Evolution, and Structure* (Columbia Univ. Press, 2004).
- Martin, T. Dryolestidae (Dryolestoidae, Mammalia) aus dem Oberen Jura von Portugal. *Abh. Senckenb. Naturforsch. Ges.* **550**, 1–119 (1999).
- Bonaparte, J. F. Sobre *Mesungulatum housayi* y nuevos mamíferos cretácicos de Patagonia, Argentina. *Actas Congr. Argent. Paleontol. Estratigr.* **2**, 63–95 (1986).

- Bonaparte, J. F. New Late Cretaceous mammals from the Los Alamitos Formation, southern Patagonia. *Nat. Geogr. Res.* **6**, 63–93 (1990).
- Bonaparte, J. F. Approach to the significance of the Late Cretaceous mammals of South America. *Berliner Geowiss.* **13**, 30–44 (1994).
- Gelfo, J. N. & Pascual, R. *Peligrotherium tropicalis* (Mammalia, Dryolestida) from the early Paleocene of Patagonia, a survival from a Mesozoic Gondwanan radiation. *Geodiversitas* **23**, 69–379 (2001).
- Rougier, G. W., Forasiepi, A. M., Hill, R. V. & Novacek, M. J. New mammalian remains from the Late Cretaceous La Colonia Formation, Patagonia, Argentina. *Acta Palaeontol. Pol.* **54**, 195–212 (2009).
- Chornogubsky, L. New remains of the dryolestoid mammal *Leonardus cuspidatus* from the Los Alamitos Formation (Late Cretaceous, Argentina). *Paläont. Z.* **85**, 1–8 (2011).
- Leanza, H. A., Apesteguía, S., Novas, F. E. & de la Fuente, M. S. Cretaceous terrestrial beds from the Neuquén Basin (Argentina) and their tetrapod assemblages. *Cretac. Res.* **25**, 61–87 (2004).
- Apesteguía, S. & Novas, F. E. Large Cretaceous sphenodontian from Patagonia provides insight into lepidosaur evolution in Gondwana. *Nature* **425**, 609–612 (2003).

12. Apesteguía, S. & Zaher, H. A. Cretaceous terrestrial snake with robust hindlimbs and a sacrum. *Nature* **440**, 1037–1040 (2006).
13. Gallina, P. A. & Apesteguía, S. *Cathartesaura anaerobica* gen. et sp. nov., a new rebbachisaurid (Dinosauria, Sauropoda) from the Huincul Formation (Upper Cretaceous), Río Negro, Argentina. *Rev. Mus. Argent. Cienc. Nat.* **7**, 153–166 (2005).
14. Makovicky, P., Apesteguía, S. & Agnolín, F. L. The earliest dromaeosaurid theropod from South America. *Nature* **437**, 1007–1011 (2005).
15. Páez Arango, N. *Dental and Craniomandibular Anatomy of Peligrotherium tropicalis: the Evolutionary Radiation of South American Dryolestoid Mammals*. Thesis, Univ. Louisville (2008).
16. Krebs, B. *Drescheratherium acutum* gen. et sp. nov., ein neuer Eupantotherier (Mammalia) aus dem Oberen Jura von Portugal. *Berliner Geowiss.* **28**, 91–111 (1998).
17. Bonaparte, J. F. & Migale, L. A. *Protomamíferos y mamíferos Mesozoicos de América del Sur* (Museo de Ciencias Naturales Carlos Ameghino, 2010).
18. Bonaparte, J. F. Una nueva especie de Triconodonta (Mammalia) de la Formación Los Alamitos, Provincia de Río Negro y comentarios sobre su fauna de mamíferos. *Ameghiniana* **29**, 99–110 (1992).
19. Rougier, G. W. & Wible, J. R. In *Amniote Paleobiology: Phylogenetic and Functional Perspectives on the Evolution of Mammals, Birds and Reptiles* (eds Carrano, M. T., Gaudin, T. J., Blob, R. & Wible, J. R.) 269–311 (Univ. Chicago Press, 2006).
20. Turnbull, W. D. Mammalian masticatory apparatus. *Fieldiana Geol.* **18**, 153–356 (1970).
21. Meng, J., Wang, Y. & Li, C. Transitional mammalian middle ear from a new Cretaceous Jehol eutriconodont. *Nature* **472**, 181–185 (2011).
22. Wible, J. R., Rougier, G. W., Novacek, M. J. & Asher, R. J. The eutherian mammal *Maelestes gobiensis* from the Late Cretaceous of Mongolia and the phylogeny of Cretaceous eutherians. *Bull. Am. Mus. Nat. Hist.* **327**, 1–123 (2009).
23. Rougier, G. W. et al. In *Dinosaurios y Paleontología desde América Latina* (eds Calvo, J., Porfiri, J., González Riga, B. & Dos Santos, D.) 195–214 (Editorial de la Universidad Nacional de Cuyo, 2011).
24. Luo, Z. X., Kielan-Jaworowska, Z. & Cifelli, R. L. In quest for a phylogeny of Mesozoic mammals. *Acta Palaeontol. Pol.* **47**, 1–78 (2002).
25. Rowe, T. B., Rich, T. H., Vickers-Rich, P., Springer, M. & Woodburne, M. O. The oldest platypus and its bearing on divergence timing of the platypus and echidna clades. *Proc. Natl Acad. Sci. USA* **105**, 1238–1242 (2008).
26. Sigogneau-Russell, D. Further data and reflexions on the tribosphenid mammals (Tribotheria) from the Early Cretaceous of Morocco. *Bull. Mus. Natl Hist. Nat. Paris* **16**, 291–312 (1995).
27. Wilson, G. P., Das Sarma, D. C. & Anantharaman, S. Late Cretaceous sudamericid gondwanatherian mammals from India with paleobiogeographic considerations of Gondwanan mammals. *J. Vertebr. Paleontol.* **27**, 521–531 (2007).
28. Gurovich, Y. & Beck, R. The phylogenetic position of the enigmatic mammalian clade Gondwanatheria. *J. Mamm. Evol.* **16**, 25–49 (2009).
29. Sigogneau-Russell, D. & Ensom, P. C. *Thereuodon* (Theria, Symmetrodonta) from the Lower Cretaceous of North Africa and Europe, and a brief review of symmetrodonts. *Cretac. Res.* **19**, 445–470 (1998).
30. Fox, R. C. Upper molar structure in the Late Cretaceous symmetrodont *Symmetrodontoides* Fox, and a classification of the Symmetrodonta (Mammalia). *J. Paleontol.* **59**, 21–26 (1985).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Salinas, P. A. Gallina and P. J. Makovicky for finding the best specimens; the Avelás and Pincheira families for logistical support; C. Muñoz and R. Barbieri from the MPCA for access to collections under their care; J. A. González and A. Davidson for illustration and technical assistance. Field work permits and loans were facilitated by M. Solorza. C. Corbitt and J. R. Wible read an earlier version of the manuscript. Field work and research was supported by the Antorchas Foundation, American Museum of Natural History and NSF grants DEB 0946430, DEB 1068089 and ATOL 0629959 (to G.W.R.), The Jurassic Foundation (to S.A.) and NASA and Field Museum Womens' Board (to P. J. Makovicky). This is L.C.G.'s R-46 contribution to the IDEAN.

Author Contributions G.W.R. wrote the manuscript with contributions from all authors, edited the figures, scored the matrix and performed the phylogenetic analysis; S.A. edited the manuscript and figures; L.C.G. edited the manuscript, figures, matrix and performed the phylogenetic analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.W.R. (grougier@louisville.edu).

Grid cells without theta oscillations in the entorhinal cortex of bats

Michael M. Yartsev¹, Menno P. Witter² & Nachum Ulanovsky¹

Grid cells provide a neural representation of space, by discharging when an animal traverses through the vertices of a periodic hexagonal grid spanning the environment¹. Although grid cells have been characterized in detail in rats^{1–6}, the fundamental question of what neural dynamics give rise to the grid structure remains unresolved. Two competing classes of models were proposed: network models, based on attractor dynamics^{7–9}, and oscillatory interference models, which propose that interference between somatic and dendritic theta-band oscillations (4–10 Hz) in single neurons transforms a temporal oscillation into a spatially periodic grid^{10–13}. So far, these models could not be dissociated experimentally, because rodent grid cells always co-exist with continuous theta oscillations^{4–6,14}. Here we used a novel animal model, the Egyptian fruit bat^{15,16}, to refute the proposed causal link between grids and theta oscillations. On the basis of our previous finding from bat hippocampus, of spatially tuned place cells in the absence of continuous theta oscillations¹⁷, we hypothesized that grid cells in bat medial entorhinal cortex might also exist without theta oscillations. Indeed, we found grid cells in bat medial entorhinal cortex that shared remarkable similarities to rodent grid cells. Notably, the grids existed in the absence of continuous theta-band oscillations, and with almost no theta modulation of grid-cell spiking—both of which are essential prerequisites of the oscillatory interference models. Our results provide a direct demonstration of grid cells in a non-rodent species. Furthermore, they strongly argue against a major class of computational models of grid cells.

To elucidate the cellular and network mechanisms of grid cells in mammalian entorhinal cortex, we conducted electrophysiological recordings in a megabat, *Rousettus aegyptiacus* (Egyptian fruit bat). This bat species¹⁵ belongs to a different suborder of bats than the big brown bat¹⁶, the hippocampus of which we studied previously^{17,18}. Therefore, to set the foundations for subsequent entorhinal recordings in this bat species, we first examined neural activity in hippocampal area CA1 (Fig. 1a). Bats crawled in a large arena (Fig. 1b) in search of food, and the activity of individual neurons was recorded using tetrodes (Supplementary Figs 1 and 2). Of all the well-isolated excitatory neurons in CA1, 36% were classified as place cells, becoming active when the bat entered a particular region of the environment (23 out of 64 cells with spatial information >0.5 bits per spike). Locations of individual place fields spanned the entire experimental arena (Fig. 1c), similar to place cells in rats¹⁹. High-frequency ripple oscillations (120–160 Hz) were present in the local field potential (LFP) during sleep (Supplementary Fig. 3a), and were very similar in their properties to ripples in rat hippocampus²⁰ (Supplementary Fig. 3 and Supplementary Text). Hippocampal theta oscillations in the LFP occurred in short, intermittent bouts (Fig. 1d), lasting typically ~1 s, and these bouts in the behaving bat were separated with an average 19-s interval (Supplementary Fig. 4a–d)—markedly different from the continuous theta oscillations observed in locomoting rats²¹, but similar to the intermittent theta bouts in monkeys²² and humans²³. Because of the sparse occurrence of these bouts, theta oscillations were not evident in

spectral analysis of the LFP, neither during sleep (Fig. 1e, left), during behavioural sessions (right), nor as a function of the bat's velocity (Fig. 1f). Thus, in agreement with our previous findings in big brown bats¹⁷, we found that in Egyptian fruit bats, hippocampal place cells existed in the absence of continuous theta oscillations in the LFP.

We next turned to studying the medial entorhinal cortex (MEC) of bats, the region where grid cells are most prevalent in rats^{1–6}. As a first step, we determined the anatomical location and borders of MEC (Fig. 2a and Supplementary Figs 5 and 6). Immunohistochemical staining revealed that: (1) similar to rats²⁴, the dorsal border of bat MEC (postrhinal border) was clearly identifiable by an abrupt change in layer structure (Supplementary Fig. 6b, d) and an abrupt transition in parvalbumin and calretinin staining (Fig. 2a and Supplementary Figs 5 and 6); and (2) similar to rats, the bat MEC had a densely packed layer II, sparser layers III and V, and a layer IV (*lamina dissecans*) with very few cell bodies (Supplementary Figs 5 and 6). Thus, the overall structure of bat MEC was very similar to that of rat MEC.

We targeted our tetrode recordings to the dorsal-most part of MEC: the region where we expected to find grid cells with the tightest and most pronounced grid structure, as in rats^{1,2}. The positions of all

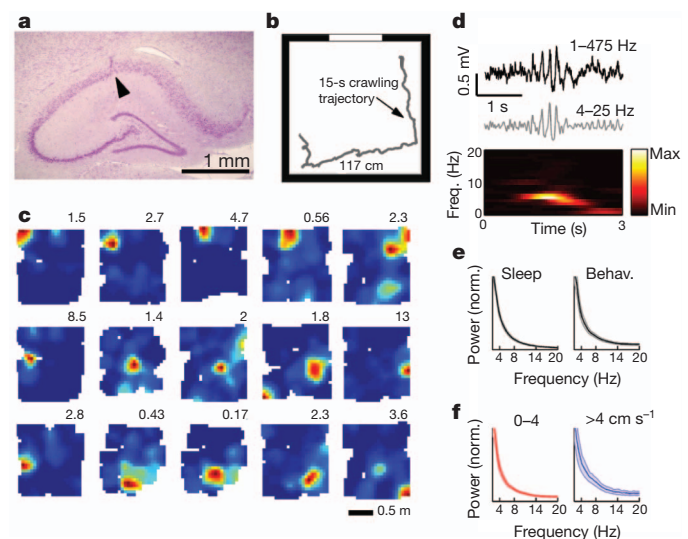


Figure 1 | Bat CA1 recordings demonstrate place cells but no continuous theta oscillations. **a**, Coronal section showing tetrode track in dorsal hippocampal area CA1 (arrowhead). **b**, Behavioural arena (117 × 117 cm) with white cue card. Grey line shows example 15-s crawling trajectory. **c**, Firing rate maps of 15 place cells, colour-coded from zero (blue) to maximum firing rate (red; value indicated). **d**, Example theta bout recorded during locomotion. Top: wideband LFP trace (1–475 Hz, black); middle, filtered trace (4–25 Hz, grey); bottom, spectrogram (frequency × time). **e**, LFP power spectrum during sleep (left) and behaviour (right) for all recording sites (mean ± s.e.m.). **f**, LFP power spectrum separated according to bat's velocity (indicated); normalized to power at 2 Hz. Note absence of peak at theta frequency.

¹Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel. ²Kavli Institute for Systems Neuroscience and Centre for the Biology of Memory, Norwegian University of Science and Technology, NO-7489 Trondheim, Norway.

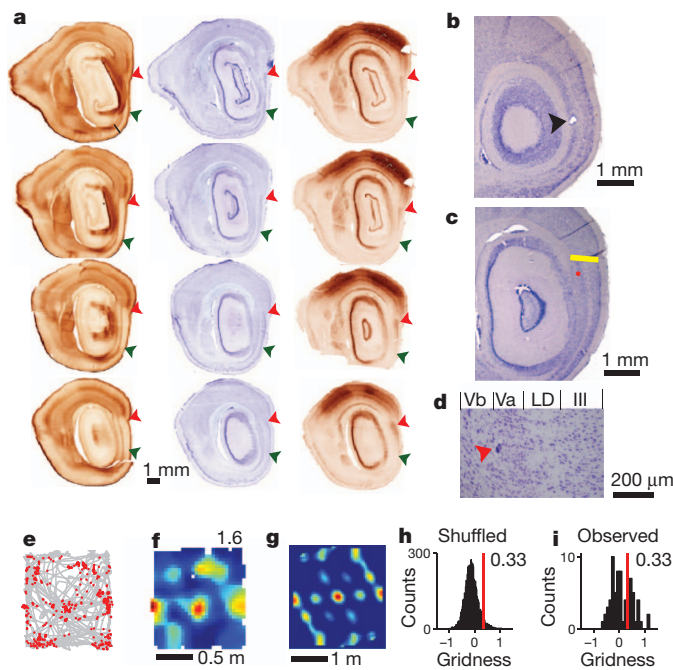


Figure 2 | Anatomical delineation of MEC in Egyptian fruit bats, and targeting of grid cells. **a**, Sagittal sections stained for calretinin (left), Nissl (middle) and parvalbumin (right), arranged from medial (top) to lateral (bottom). Arrowheads indicate MEC borders: dorsal, red; ventral, green. **b**, Electrolytic lesion at the end of a tetrode track (arrowhead). **c**, Recording location in MEC layer V (red dot), close to the dorsal (postrhinal) border of MEC (yellow line). **d**, Higher magnification of the recording location in **c** (arrowhead). MEC layers are indicated on top; LD, lamina dissecans²⁴ (layer IV). **e–g**, Grid cell recorded in the MEC location shown in **c**. **e**, Trajectory (grey line) and locations of spikes (red dots), showing the hexagonal firing pattern of a grid cell. **f**, Colour-coded rate map, with peak rate indicated. **g**, Spatial autocorrelogram of the rate map. **h**, Distribution of gridness for shuffled data (Supplementary Information), with 95th percentile indicated (red line; gridness = 0.33). **i**, Histogram of gridness values for all well-separated MEC neurons ($n = 70$). Neurons with gridness > 0.33 were defined as grid cells ($n = 25$).

tetrodes were verified histologically by reconstructing the tetrode tracks and by electrolytic lesions (Fig. 2b–d and Supplementary Figs 7–9). Many MEC neurons were active at multiple locations in the arena (for example, Fig. 2e, f), and a hexagonal grid structure was revealed by computing the spatial autocorrelogram of their firing-rate map (Fig. 2g). To quantify the degree of 60° hexagonal rotational symmetry for each neuron, we employed a commonly used ‘gridness’ index² (Supplementary Fig. 10): a higher index indicates a more hexagonal firing field. To test statistically whether a neuron is a grid cell, we used a standard shuffling procedure^{4–6} (Fig. 2h, i). Of the 70 well-isolated, behaviourally active neurons recorded in MEC, 36% were classified as grid cells (25 out of 70).

How similar are grid cells in bats versus rats? First, similar to rats¹, grid vertices were separated by 60° angles (Fig. 3a), as expected from a hexagonal structure, and individual firing fields were almost equally spaced (mean standard deviation of grid spacing across the six inner peaks: 6.8 cm; $n = 25$ grid cells). Second, as in rats^{1–6}, co-localized grid cells in superficial layers of MEC shared similar grid orientation (tilt) and spacing (Fig. 3b and Supplementary Information). Third, as in rats¹, cross-correlating the rate maps of simultaneously recorded grid cells revealed that their maps were offset in phase (Fig. 3c, central peaks of cross-correlograms are offset from the white cross), and the amount of offset spanned all possible phases (Fig. 3d). Fourth, as in rats¹, the grid spacing of individual cells increased with the cell’s distance from the postrhinal border (Fig. 3e); this correlation was significant across all neurons ($r = 0.47$, $P < 0.025$), but was even more apparent when

analysing separately each bat’s data (bat 1, $r = 0.98$, $P < 0.02$; bat 2, $r = 0.67$, $P < 0.05$; bat 3, $r = 0.60$, $P = 0.07$). Fifth, as in rats², the firing rate increased with movement velocity (Fig. 3f; t -test, $t_{24} = 2.9$, $P < 0.001$). This correlation of firing rate and velocity probably explains the relatively low firing rates of MEC neurons in crawling bats (Fig. 3e, g) compared to rats^{1–3,6}, because bats crawled rather slowly (Supplementary Fig. 11). Finally, we found that in bat MEC, the diversity of spatial cell types was very similar to that known from rat MEC^{1–6,25} and contained: (1) ‘pure grid cells’^{1,2} (Fig. 3g, lower-right quadrant; for example, cells 3, 4, 5), with no sensitivity to the animal’s head direction^{1,2}; (2) ‘conjunctive grid cells’^{2,6} (Fig. 3g, upper-right quadrant; cell 2), which are grid cells tuned to a specific head direction^{2,6}; (3) head-direction cells² that have no grid structure (Fig. 3g, upper-left quadrant; cells 8, 9; see also Supplementary Fig. 12); and (4) ‘border cells’²⁵, which fired along geometrical borders of the environment (for example, Fig. 3g, cell 6). Some cells that did not cross the shuffling thresholds still exhibited clear spatial patterns characteristic of grid cells (Fig. 3g, cell 10) or head-direction cells (cell 7). Taken together, these results indicate that the detailed properties of grid cells (and other cell types) in bat MEC were very similar to those in rat MEC.

We next turned to address the central question of this study, and asked whether the grids can exist without key elements of the oscillatory interference models^{10–13}—namely, without continuous theta oscillations in the LFP, and in the absence of theta modulation of grid-cell firing. As a first step in studying the LFP in MEC, we examined non-theta-band, high-frequency ripple oscillations, which are most prominent during sleep²⁰ (Supplementary Fig. 3d). Ripples in bat MEC were very similar in their properties to ripples in rat MEC (Supplementary Fig. 3d–f and Supplementary Text). However, despite the similarity in ripple oscillations, bats differed markedly from rats in the nature of theta oscillations. First, unlike rats^{4–6,14,26}, we never observed a prominent continuous theta-band oscillation in the LFP, irrespective of the recording site in MEC and the type of reference used (Fig. 4a, b and Supplementary Information). The LFP power-spectrum showed no theta peak, neither during sleep (Supplementary Fig. 13, left), during behaviour (Fig. 4a, b, ‘Behav.’), as a function of the animal’s velocity (Fig. 4a, b, coloured panels), nor as a function of the bat’s echolocation mode¹⁷ (Supplementary Fig. 14a–d). Second, unlike the continuous theta oscillations typically observed in rat MEC^{4–6,14,26}, theta oscillations in bat MEC occurred in short intermittent bouts, both during behaviour (Fig. 4c–e and Supplementary Fig. 4f, h) and during sleep (Supplementary Fig. 4e, g), similar to bat CA1; 92% of theta bouts during behaviour had duration ≤ 1 s (Fig. 4d). Theta bouts were separated by very long inter-bout intervals (Fig. 4e; average interval: 37 ± 2 s; 18% of intervals were > 1 min). The bat’s velocity and echolocation rate were not different during theta bouts versus non-theta epochs (Supplementary Fig. 14e, f). Third, we computed the spike-train temporal autocorrelations for individual grid cells (Fig. 4f, top and Supplementary Fig. 15), and examined the degree of theta modulation in these autocorrelations by computing a standard ‘theta index’ that was used in previous studies in rats^{4–6,26}; this index is based on the relative theta-band power in the Fourier transform of the temporal autocorrelation (Fig. 4g and Supplementary Information). Using either the same criterion that was employed previously to identify theta modulation in rat grid cells⁶ (theta index > 5), or using a statistical shuffling procedure for each individual spike train, we found that nearly all grid cells in bat MEC (24 out of 25) did not exhibit theta-modulated firing (Fig. 4h, left; theta index across all grid cells 1.29 ± 0.82 (mean \pm s.d., maximal value 2.9)), irrespective of the bat’s velocity (Supplementary Fig. 16). Because, in the rat MEC, neurons recorded in layers II and III show the most pronounced theta modulation of neuronal firing^{6,27}, we also analysed separately the data recorded from layers II and III of bat MEC; these analyses showed that no significant theta modulation was present in any of the 35 neurons (of all classes) recorded in layers II and III and in any of the multi-units from these layers (Supplementary Fig. 17). Fourth, because the firing of most neurons in rat MEC (especially in superficial layers) is

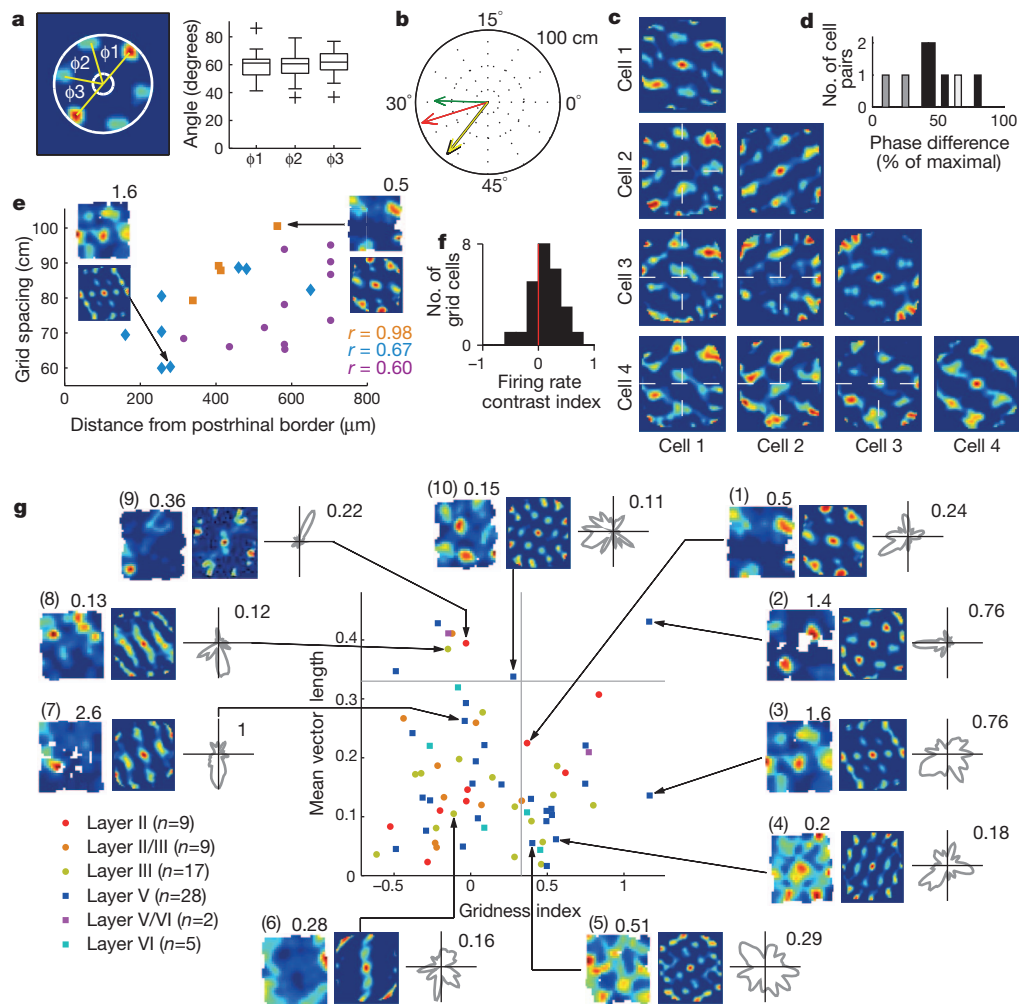


Figure 3 | Grid-cell properties in the bat resemble those in the rat. **a**, Left: definition of the three hexagon angles, depicted on the autocorrelogram annulus of a grid cell. Right: distribution of angles ($n = 25$ grid cells). Box plots show the median angles and interquartile and total ranges; +, outliers. **b**, Polar plot (60° cycle) showing four co-localized grid cells recorded simultaneously on the same tetrode in the superficial layers of MEC, depicting grid spacing (arrow length, cm) and orientation (angle within 60° cycle). Red, green, black and yellow arrows denote these four co-localized grid cells. **c**, Autocorrelograms (diagonal panels) and cross-correlograms (off-diagonal) of the four simultaneously recorded grid cells from **b**. **d**, Distribution of phase differences between pairs of co-localized grid cells. Different shading indicates different recording sites. **e**, Grid spacing increases with distance from postrhinal (dorsal)

border of MEC. Colours of dots indicate individual bats and match the colours of r values. Insets show firing-rate maps and autocorrelograms of two grid cells recorded at different distances from the postrhinal border. **f**, Distribution of firing-rate contrast indexes for high versus low movement velocity. **g**, Scatter plot of mean vector length index versus gridness index for all neurons from MEC ($n = 70$). Colour and shape of symbols denote recording layer in MEC. Vertical and horizontal lines indicate 95th percentile of shuffled distributions of gridness and mean vector length, respectively. Insets show examples of neurons discussed in the main text, showing (from left to right) the rate map, spatial autocorrelogram and head-direction polar rate plots (grey). Peak firing rate (in Hz) is indicated for rate maps and polar maps.

theta-modulated⁶, and they have similar phases²⁸, we also examined the temporal periodicity of the multi-unit activity, where firing rates are much higher than in individual neurons, and hence oscillations might be detected more robustly. We found that 100% of multi-unit sites where grid cells were recorded (17 out of 17) did not exhibit theta-modulated firing (Fig. 4h, right; theta index 0.97 ± 0.56 ; maximal value 2.1; see also Supplementary Figs 15–17). Fifth, because rat MEC neurons are often locked to a specific phase of the theta cycle^{14,27,28}, we examined whether spikes are locked to bats' theta phase during theta bouts. We found that bat MEC neurons indeed exhibited a clear, albeit weak, phase-locking during theta bouts (Fig. 4i, bottom, grey; see also Supplementary Fig. 18). Importantly, no phase locking could be observed outside the theta bouts (Fig. 4i, red). The contrast between phase locking of spikes during theta bouts versus lack of locking outside the bouts (Fig. 4i) indicates that theta bouts in bats are truly discrete and locally generated events. Sixth, to examine the possible contribution of theta bouts to grid formation, we removed in each grid cell all theta-bout

epochs, and re-computed the firing-rate maps and two-dimensional autocorrelograms; this did not cause substantial alterations in the grid pattern (Fig. 4j, k and Supplementary Fig. 19). Notably, only a minority of spikes emitted by any single grid cell occurred during theta bouts ($4.4 \pm 0.75\%$). In fact, in some grid cells, the grid field existed in the absence of any spikes emitted during theta bouts (Fig. 4k, right; zero spikes emitted during all the theta bouts). Population analysis confirmed that theta-bout removal did not lead to significant changes in gridness values, in any of the grid cells (Fig. 4l, 100% of the cells showed changes in gridness that did not exceed the 95% confidence intervals). This suggests that the grids are maintained during times when theta oscillations are not present.

Taken together, these findings provide the first report on grid cells in a non-rodent species, which supports the generality of the grid-cell phenomenon across mammals; more importantly, our findings causally dissociate the link between the existence of grids and the existence of continuous theta-band oscillations in the mammalian entorhinal

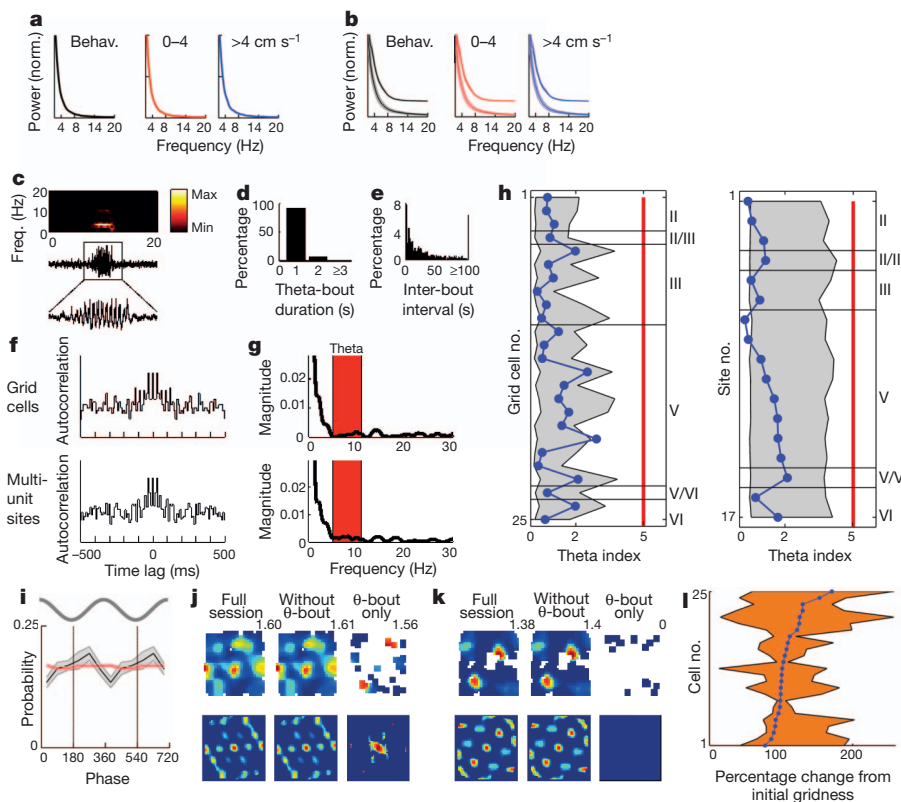


Figure 4 | No continuous theta oscillations in bat MEC. **a**, LFP power spectrum during a single behavioural session (left, Behav.), and separated according to bat's velocity (coloured panels). **b**, Population average LFP power spectrum for all behavioural sessions in which grid cells were recorded (left, Behav.), and separated by velocity (coloured panels). The two lines in each panel show the two reference types used for electrophysiological recordings: single-ended (ground screw, top line) and double-ended (local reference tetrode, bottom); vertical separation is for display purposes only (mean \pm s.e.m.). Note the absence of a peak at theta frequency. **c**, LFP trace showing an example theta bout during locomotion. Middle 20-s wideband trace (1–475 Hz). The box around the theta bout serves as calibration bars: 7-s horizontal and 0.4-mV vertical. Bottom: 7-s zoom in. Top: spectrogram. **d**, Distribution of theta-bout durations during behaviour. **e**, Distribution of inter-bout intervals during behaviour. **f**, Spike-train temporal autocorrelograms for a single grid cell (top) and the multi-unit activity recorded at the same site (bottom). Bin size, 10 ms. **g**, Power spectrum of the autocorrelograms in **f**. Note the absence of spectral peak at the theta frequency

cortex. The similarities in the anatomy and grid-cell properties between bats and rats strongly suggest that similar underlying neural mechanisms generate the grid, indicating that the functional dissociation between theta and grids generalizes across mammalian species. Although it is possible that continuous theta oscillations would be found in bats during flight, we emphasize that, under the specific crawling conditions of our experiments, we observed simultaneously the existence of grids without continuous theta oscillations in the LFP or in the spiking activity, which strongly argues against the theta-based class of computational models of grid cells^{10–13}, but is consistent with the other models which do not rely on theta oscillations^{7–9}.

Recently, two studies showed that inactivation of medial septum inputs to rat MEC disrupts grid fields, and also disrupts theta oscillations in MEC^{28,29}. Additionally, these studies found that, after medial septum inactivation, the firing rates of MEC grid cells dropped by >40%, on average, and in many cases the firing rates dropped three-fold and even fivefold^{28,29}. These observations have been interpreted as supporting the oscillatory interference models^{28,29}, but they equally well support network models of grid cells: it is well known in the theory of neural networks that the removal of a major input to a network, if accompanied by a marked decrease in firing rates, can drive the

range (5–11 Hz, highlighted in red). **h**, Theta indexes for grid cells ($n = 25$, left) and multi-unit sites in which grid cells were recorded ($n = 17$, right), arranged by recording layer in MEC (layer number indicated on the right). Shaded area indicates 95% confidence intervals. Notably, 96% (24 out of 25) of grid cells and 100% (17 of 17) of multi-unit sites did not show significant theta modulation, irrespective of the recording layer. **i**, Population average phase histograms of MEC neurons' discharge probability during theta bouts (grey curve) and during non-theta epochs (red). Both curves show mean \pm s.e.m. Two theta cycles are shown for visualization purposes. **j**, **k**, Removal of theta bouts: rate maps (top) and spatial autocorrelograms (bottom) of two grid cells (**j** and **k**), for all the data (left), without theta bouts (middle) and during theta bouts only (right). Colour scale of rate maps normalized to maximal firing rate for all three maps for each neuron. **l**, Changes in gridness after theta-bout removal; data for all grid cells (blue dots, sorted) and 95% bootstrap confidence interval (orange area; Supplementary Information) are shown. None of the grid cells showed significant changes in gridness after removal of theta bouts.

network into a very different activity regime³⁰, which could disrupt grid formation. Thus, inactivation of a major input to a brain network^{28,29} cannot serve to dissociate oscillation-based models from network models. In contrast, our study in bats did not manipulate the inputs to the entorhinal network, and thus provided a unique opportunity to causally challenge the oscillatory interference models of grid cells. More generally, we provide here a rare example of causally disproving a major class of computational models of a higher brain area.

METHODS SUMMARY

Immunohistochemical stainings were conducted to delineate the anatomical location and borders of the medial entorhinal cortex (MEC) in Egyptian fruit bats (*Rousettus aegyptiacus*). Single neuron activity and local field potentials (LFP) were recorded from hippocampal area CA1 and MEC of five bats (two and three bats, respectively), using tetrodes^{17–19}. Neuronal activity and positional data were collected while bats foraged in a large arena (117 \times 117 cm) in search of food. Place cells¹⁹ in CA1 were identified using a criterion of spatial information >0.5 bits per spike (ref. 17). Grid cells and head-direction cells were identified using a standard gridness index² and mean vector length index of the head-direction tuning^{4,5}; significance of these two indices was tested using a random shuffling procedure similar to that described previously in rats^{4,5}. To quantify grid properties, we computed grid spacing, orientation, phase and velocity-modulation of the cell's

firing^{1,2}. High-frequency ripple oscillations in the LFP during sleep were detected as transients in the power of the filtered LFP trace (80–160 Hz) exceeding 7 s.d. above the mean power¹⁷. Theta oscillation epochs were defined as 2-s windows in which the ratio between the power in the theta (4–8 Hz) and delta (2–4 Hz) frequency ranges exceeded 2.0. Theta modulation of neuronal firing was assessed using a standard theta index, which is based on the spectral power of the spike train temporal autocorrelogram^{4–6}. Detailed experimental and analytical procedures are provided in the Supplementary Information.

Received 9 April; accepted 27 September 2011.

- Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005).
- Sargolini, F. *et al.* Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science* **312**, 758–762 (2006).
- Barry, C., Hayman, R., Burgess, N. & Jeffery, K. J. Experience-dependent rescaling of entorhinal grids. *Nature Neurosci.* **10**, 682–684 (2007).
- Langston, R. F. *et al.* Development of the spatial representation system in the rat. *Science* **328**, 1576–1580 (2010).
- Wills, T. J., Cacucci, F., Burgess, N. & O'Keefe, J. Development of the hippocampal cognitive map in preweanling rats. *Science* **328**, 1573–1576 (2010).
- Boccara, C. N. *et al.* Grid cells in pre- and parasubiculum. *Nature Neurosci.* **13**, 987–994 (2010).
- Fuhs, M. C. & Touretzky, D. S. A spin glass model of path integration in rat medial entorhinal cortex. *J. Neurosci.* **26**, 4266–4276 (2006).
- McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. & Moser, M.-B. Path integration and the neural basis of the 'cognitive map'. *Nature Rev. Neurosci.* **7**, 663–678 (2006).
- Burak, Y. & Fiete, I. R. Accurate path integration in continuous attractor network models of grid cells. *PLOS Comput. Biol.* **5**, e1000291 (2009).
- Burgess, N., Barry, C. & O'Keefe, J. An oscillatory interference model of grid cell firing. *Hippocampus* **17**, 801–812 (2007).
- Giocomo, L. M., Zilli, E. A., Fransen, E. & Hasselmo, M. E. Temporal frequency of subthreshold oscillations scales with entorhinal grid cell field spacing. *Science* **315**, 1719–1722 (2007).
- Hasselmo, M. E., Giocomo, L. M. & Zilli, E. A. Grid cell firing may arise from interference of theta frequency membrane potential oscillations in single neurons. *Hippocampus* **17**, 1252–1271 (2007).
- Blair, H. T., Welday, A. C. & Zhang, K. Scale-invariant memory representations emerge from Moiré interference between grid fields that produce theta oscillations: a computational model. *J. Neurosci.* **27**, 3211–3229 (2007).
- Hafting, T., Fyhn, M., Bonnevie, T., Moser, M.-B. & Moser, E. I. Hippocampus-independent phase precession in entorhinal grid cells. *Nature* **453**, 1248–1252 (2008).
- Yovel, Y., Falk, B., Moss, C. F. & Ulanovsky, N. Optimal localization by pointing off axis. *Science* **327**, 701–704 (2010).
- Ulanovsky, N. & Moss, C. F. What the bat's voice tells the bat's brain. *Proc. Natl Acad. Sci. USA* **105**, 8491–8498 (2008).
- Ulanovsky, N. & Moss, C. F. Hippocampal cellular and network activity in freely moving echolocating bats. *Nature Neurosci.* **10**, 224–233 (2007).
- Ulanovsky, N. & Moss, C. F. Dynamics of hippocampal spatial representation in echolocating bats. *Hippocampus* **21**, 150–161 (2011).
- Wilson, M. A. & McNaughton, B. L. Dynamics of the hippocampal ensemble code for space. *Science* **261**, 1055–1058 (1993).
- Chrobak, J. J. & Buzsáki, G. High-frequency oscillations in the output networks of the hippocampal-entorhinal axis of the freely behaving rat. *J. Neurosci.* **16**, 3056–3066 (1996).
- Buzsáki, G. Theta oscillations in the hippocampus. *Neuron* **33**, 325–340 (2002).
- Stewart, M. & Fox, S. E. Hippocampal theta activity in monkeys. *Brain Res.* **538**, 59–63 (1991).
- Ekstrom, A. D. *et al.* Human hippocampal theta activity during virtual navigation. *Hippocampus* **15**, 881–889 (2005).
- Witter, M. P., Wouterlood, F. G., Naber, P. A. & Van Haeften, T. Anatomical organization of the parahippocampal-hippocampal network. *Ann. NY Acad. Sci.* **911**, 1–24 (2000).
- Solstad, T., Boccarra, C. N., Kropff, E., Moser, M.-B. & Moser, E. I. Representation of geometric borders in the entorhinal cortex. *Science* **322**, 1865–1868 (2008).
- Deshmukh, S. S., Yoganarasimha, D., Voicu, H. & Knierim, J. J. Theta modulation in the medial and the lateral entorhinal cortices. *J. Neurophysiol.* **104**, 994–1006 (2010).
- Mizuseki, K., Sirota, A., Pastalkova, E. & Buzsáki, G. Theta oscillations provide temporal windows for local circuit computation in the entorhinal-hippocampal loop. *Neuron* **64**, 267–280 (2009).
- Brandon, M. P. *et al.* Reduction of theta rhythm dissociates grid cell spatial periodicity from directional tuning. *Science* **332**, 595–599 (2011).
- Koenig, J., Linder, A. N., Leutgeb, J. K. & Leutgeb, S. The spatial periodicity of grid cells is not sustained during reduced theta oscillations. *Science* **332**, 592–595 (2011).
- Hansel, D. & Sompolinsky, H. Modeling feature selectivity in local cortical circuits. In *Methods in Neuronal Modeling: from Synapses to Networks* 2nd edn (eds Kock, C. & Segev, I.) 499–567 (MIT Press, 1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Derdikman, S. Romani, M. Ahrens and Y. Cohen for comments on the manuscript, M. Melcón for initial assistance with hippocampal CA1 recordings, M. Weinberg for veterinary oversight, and R. Eilam and C. Ra'anani for histology. This study was supported by research grants from the Israel Science Foundation and Minerva Foundation to N.U., by a Lev-Zion predoctoral excellence fellowship to M.M.Y., as well as by grants from the Norwegian Research Council and the Kavli Foundation to M.P.W.

Author Contributions M.M.Y. and N.U. designed the study, conducted the experiments, analysed the data and wrote the manuscript. M.P.W. performed immunohistochemical analyses, delineated the anatomical structures in the hippocampus and entorhinal cortex of Egyptian fruit bats, and verified recording sites. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.U. (nachum.ulanovsky@weizmann.ac.il).

Feedback from rhodopsin controls *rhodopsin* exclusion in *Drosophila* photoreceptors

Daniel Vasilias¹, Esteban O. Mazzoni^{1†}, Simon G. Sprecher^{1†}, Konstantin Brodetskiy¹, Robert J. Johnston Jr¹, Preetmoninder Lidder¹, Nina Vogt¹, Arzu Celik^{1†} & Claude Desplan¹

Sensory systems with high discriminatory power use neurons that express only one of several alternative sensory receptor proteins. This exclusive receptor gene expression restricts the sensitivity spectrum of neurons and is coordinated with the choice of their synaptic targets^{1–3}. However, little is known about how it is maintained throughout the life of a neuron. Here we show that the green-light sensing receptor rhodopsin 6 (Rh6) acts to exclude an alternative blue-sensitive rhodopsin 5 (Rh5) from a subset of *Drosophila* R8 photoreceptor neurons⁴. Loss of Rh6 leads to a gradual expansion of Rh5 expression into all R8 photoreceptors of the ageing adult retina. The Rh6 feedback signal results in repression of the *rh5* promoter and can be mimicked by other *Drosophila* rhodopsins; it is partly dependent on activation of rhodopsin by light, and relies on G_{αq} activity, but not on the subsequent steps of the phototransduction cascade⁵. Our observations reveal a thus far unappreciated spectral plasticity of R8 photoreceptors, and identify rhodopsin feedback as an exclusion mechanism.

In the *Drosophila* visual system, rhodopsins (Rh), G-protein-coupled receptors, detect light and initiate the phototransduction cascade leading to depolarization of photoreceptor neurons⁵. Each ommatidium, the unit eye of the adult retina, contains eight photoreceptors. Six outer photoreceptors, R1–R6, express Rh1 and are involved in motion detection and dim light vision (reviewed in ref. 4). Inner photoreceptors R7 and R8 mediate colour vision and define two main ommatidial subtypes based on the rhodopsins they express: in pale (p) ommatidia, pR7 expresses ultraviolet-sensitive Rh3 whereas pR8 expresses Rh5; in yellow (y) ommatidia, yR7 expresses a distinct ultraviolet-sensitive Rh4 whereas yR8 expresses Rh6⁴. Subtypes p and y are distributed stochastically throughout the main part of the retina with an approximate 30:70 ratio (Fig. 1c)⁶. An exception to the exclusive rhodopsin expression exists in the medio-dorsal area of the eye, where although the p and y subsets are correctly specified, Rh3/Rh4 are co-expressed in yR7s⁷. This rhodopsin expression pattern is established by a well-understood developmental program executed during pupal stages^{4,8,9} (Supplementary Fig. 1a, b).

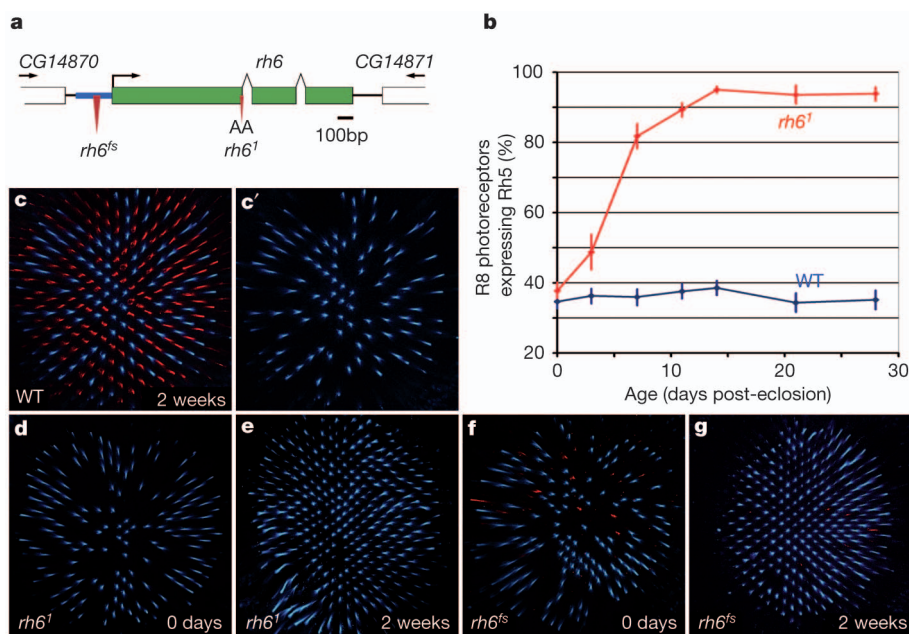


Figure 1 | Rh6 acts to repress Rh5 expression in yR8 photoreceptors.

a, Genomic *rh6* locus. The promoter region sufficient to drive *rh6* expression in yR8 is in blue, exons are in green and mutations in red. In *rh6^{fs}* mutants, 58 bp of the promoter are deleted. In *rh6^l* mutants, 21 bp at the first exon–intron junction are replaced with AA, leading to an immediate truncation of the open reading frame. **b**, Percentage of R8 photoreceptors expressing Rh5 as a function of time (days post-eclosion) in wild-type (WT, blue) and *rh6^l* mutants (red). Error bars, 84% confidence intervals. **c–g**, Wholemount retinas stained with

specific antibodies for Rh5 (blue) and Rh6 (red). **c**, **c'**, Normal expression of Rh5 and Rh6 in 2-week-old flies. **c'**, Rh5 alone. **d**, **e**, In *rh6^l* mutants, Rh5 is gradually de-repressed. At eclosion, retinas have a normal number of Rh5-expressing R8s (**d**). By 2 weeks post-eclosion, most R8s express Rh5 (**e**). **f**, **g**, *rh6^{fs}* promoter mutation leads to loss of detectable Rh6 expression in almost all yR8s. As in *rh6^l* mutants (**d**, **e**), at eclosion *rh6^{fs}* retinas have a normal number of Rh5-expressing R8s (**f**), but by 2 weeks post-eclosion, most R8 express Rh5 (**g**).

¹Center for Developmental Genetics, Department of Biology, New York University, New York, New York 10003, USA. [†]Present addresses: Departments of Pathology, Neurology, and Neuroscience, Columbia University Medical Center, 630 West 168 Street, New York, New York 10032, USA (E.O.M.); Department of Biology, University of Fribourg, Chemin du Musée 10, 1700 Fribourg, Switzerland (S.G.S.); Department of Molecular Biology and Genetics, Bogazici University, 34342 Bebek, Istanbul, Turkey (A.C.).

It is unknown, however, how p and y photoreceptor subtypes are maintained in the adult fly. The example of vertebrate olfaction, where sensory receptors act to repress expression of alternative receptor genes^{10–14}, led us to ask whether rhodopsins themselves participate in maintaining their mutual exclusion by analysing rhodopsin expression in various *rh* mutants. We found that in *rh6¹* mutants (Fig. 1a), the number of R8 cells expressing Rh5 increases dramatically and that this expansion of Rh5 is age dependent (Fig. 1b, d, e and Supplementary Table 1). In 1-day-old *rh6¹* mutant flies, Rh5 expression appears normal, with approximately 38% of R8s expressing uniformly high levels of Rh5 protein. In 3-day-old flies, additional R8s begin to express low levels of Rh5. By 14 days, nearly all (95%) R8s express Rh5. The levels of ectopic Rh5 in individual yR8s also increase over time, but remain variable and often are lower than in pR8 (Fig. 1e). In control flies, the number of Rh5-expressing R8s and the levels of expression remain stable as the flies age (36%, Fig. 1b, c and Supplementary Table 1).

We next asked if other rhodopsins are controlled by rhodopsin-mediated repression. We examined whether Rh6 expression is de-repressed in *rh5* mutants, but did not detect any Rh6 protein in pR8 in 3-week-old *rh5* mutants (Supplementary Fig. 2f). Expression of the non-R8 rhodopsins Rh1, Rh3 and Rh4 also remains normal in *rh5* or *rh6* mutants older than 3 weeks as well as in *rh5;rh6* double mutants (Supplementary Fig. 2a–e, g, h). Nonsense mutations in *rh3* or *rh4* genes do not affect expression of the remaining rhodopsins in R7s in either young or old (over 3 weeks) flies (Supplementary Figs 3 and 4, and data not shown). Thus, a rhodopsin-dependent mechanism for controlling rhodopsin expression occurs only in yR8s. Moreover, Rh5 is the only rhodopsin that is actively repressed by Rh6.

In the *rh6¹* allele, commonly found in laboratory stocks, a short deletion that spans the first exon–intron junction leads to a truncation of the protein after its fifth transmembrane domain¹⁵ (Fig. 1a). The levels of *rh6* messenger RNA (mRNA) measured by quantitative PCR with reverse transcription (qRT–PCR) are more than tenfold lower in *rh6¹* mutants than in wild-type flies (Supplementary Fig. 9a), probably because of nonsense-mediated decay. The Rh5 de-repression phenotype does not become more severe when *rh6¹* is placed over a deficiency (Supplementary Fig. 9b), suggesting that *rh6¹* is a null allele. Also, *rh6¹* can be rescued by a 2,575 base pair (bp) genomic fragment encompassing the *rh6* locus (Supplementary Figs 7a and 9b). Hereafter we refer to both *rh6¹* homozygotes and *rh6¹* trans-heterozygotes over a deficiency as *rh6* mutants and, unless otherwise noted, all phenotypes described are in ‘old’ flies 2 weeks post-eclosion or older.

We identified a second *rh6* allele, also in a laboratory stock, which we named *frank sinatra* (*rh6^{fs}*) after the singer known as ‘Ol’ blue eyes’ (Fig. 1a). This mutation removes 58 bp of the *rh6* regulatory region without affecting the coding sequence. In *rh6^{fs}* mutants, Rh6 protein is detectable only in a few R8s in retinas of young flies ($6.5\% \pm 4.4$ SD, Supplementary Fig. 9b) where it is expressed at levels generally lower than normal (Fig. 1f, g). As in *rh6¹* mutants, Rh5 is initially expressed normally in 41% of R8 in *rh6^{fs}* flies (Fig. 1f), leaving most yR8s devoid of rhodopsin expression. However, Rh5 becomes broadly de-repressed in R8s of old flies (Fig. 1g and Supplementary Fig. 9b). Rh5 is rarely expressed in the few Rh6-positive R8s of *rh6^{fs}* mutants and co-expression only occurs in cells with low Rh6 levels (not shown). We also used a *rh6* promoter-based driver (*rh6-Gal4*) to express a RNA interference (RNAi) construct targeting *rh6* in differentiated yR8s. Although this does not completely abolish Rh6 in yR8 rhabdomeres, it leads to de-repression of Rh5 in old flies (Supplementary Fig. 5a). These results support the idea that reducing the levels of normal Rh6 activity leads over time to de-repression of Rh5 expression in yR8s.

Repression of Rh5 by Rh6 in wild-type yR8 could occur transcriptionally, or post-transcriptionally. We thus asked whether *rh5* mRNA expression is de-repressed in *rh6* mutants by performing *in situ* hybridization. *rh5* mRNA is present in many more R8s in old *rh6*

mutants than in age-matched wild-type flies (Fig. 2a, b). To visualize this phenotype more clearly, we repeated the experiment in a *sevenless* (*sev*) mutant background in which R7 photoreceptors are absent¹⁶. Because specification of *rh5*-expressing pR8s depends on the overlying pR7s (Supplementary Fig. 1a), most cells become yR8 and express Rh6 in *sev* flies while Rh5 is only expressed in a few R8 photoreceptors^{17–19} (~3%, Fig. 2c). However, in old *sev;rh6* double mutants, *rh5* mRNA is de-repressed in most R8s (Fig. 2d). We also quantified changes in *rh5* mRNA expression using qRT–PCR: in 2-week-old *rh6* mutants, *rh5* mRNA more than doubles over normal levels (Supplementary Fig. 9a). To ask whether this occurs through repression of the *rh5* promoter rather than by affecting mRNA stability, we analysed the expression of a *rh5* reporter (*rh5>GFP* (green fluorescent protein)) containing a –690 to +50 *rh5* promoter fragment²⁰. In control flies, *rh5>GFP* is co-expressed with Rh5 protein in pR8s (Fig. 2e). In *rh6* mutants *rh5>GFP* expression begins normally but with age expands to most yR8s (Fig. 2f and Supplementary Fig. 9f). This supports the model that Rh6 generates a feedback signal that acts to repress transcription from the *rh5* promoter and that the relevant regulatory sites are contained within the short promoter fragment of the *rh5>GFP* transgene.

Expression of *rh5* in yR8s of *rh6* mutants could be due to a change in yR8 cell identity, either during specification or in adults. To test this, we first asked whether a reporter for *rh6* expression (*rh6-lacZ*) is correctly activated in *rh6* mutant flies. In young *rh6* mutants, *rh6-lacZ* is robustly expressed in R8s in a pattern complementary to Rh5 expression (Fig. 3a and Supplementary Fig. 9c), suggesting correct specification of the yR8 subtype. As the fly ages, these cells de-repress Rh5 but remain positive for β -galactosidase (β Gal) (Fig. 3b and Supplementary Fig. 9c). We also tested for a possible yR8-to-pR8 fate transition using the marker genes that specify these cells. The p versus y subtype specification of R8 cells depends on an R8-intrinsic bistable

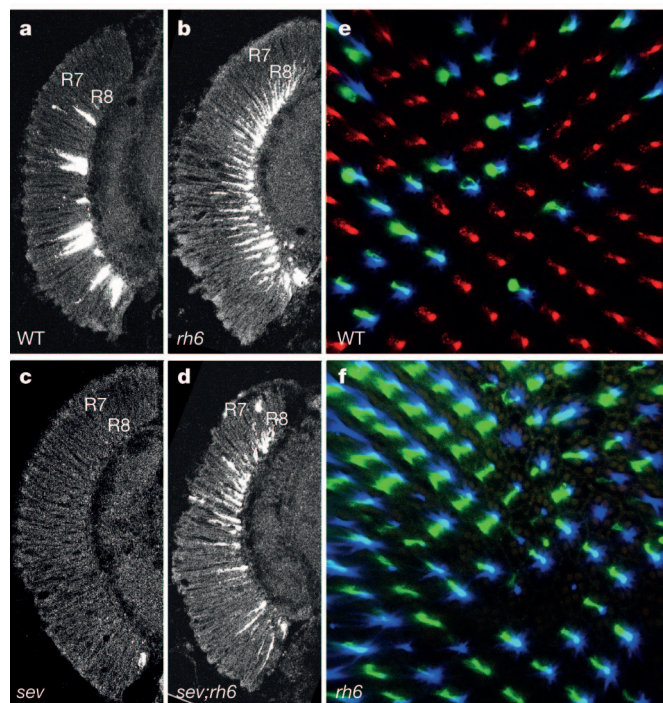


Figure 2 | Rh6 represses transcription of the *rh5* gene. a–d, *rh5* mRNA, detected by *in situ* hybridization in transverse cryo-sections of 3-week-old fly eyes. Many more cells are expressing *rh5* mRNA in the R8 layer of *rh6* mutants (b) compared with wild-type flies (a). In *sev* mutants, very few cells express *rh5* (c). However, in *sev;rh6* double mutants, *rh5* is extensively de-repressed in R8 photoreceptors (d). e, f, In 3-week-old control flies, a *rh5* reporter (*rh5>GFP*) (green) is expressed in pR8 cells that also express Rh5 protein (blue), but not in yR8 cells which express Rh6 (red) (e). In *rh6* mutants, *rh5>GFP* is de-repressed in most yR8 cells (f).

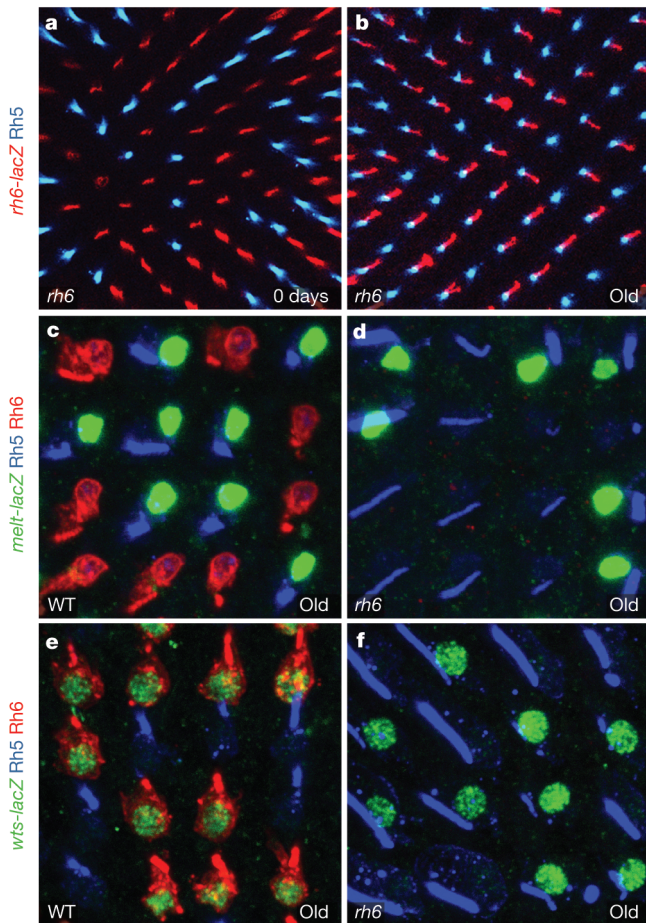


Figure 3 | Mutation of *rh6* does not lead to change in yR8 cell identity.

a, b, A *rh6-lacZ* reporter (red) is expressed normally in *rh6* mutants. It is induced in a pattern complementary to the expression of Rh5 (blue) in young flies (**a**). In 2-week-old *rh6* mutants, Rh5 expression expands into the *lacZ*-positive, yR8 cells (**b**). **c–f,** Z-projections of confocal stacks encompassing nuclei and Rh-containing rhabdomeres of R8 photoreceptors. **c, d,** Expression of the nuclear pR8 marker *melt-nlacZ* (green) does not change in *rh6* mutants. In wild-type flies it is expressed with Rh5 (blue) in pR8 and never in Rh6-expressing yR8 cells (red) (**c**). In 5-week-old *rh6* mutants, *melt-nlacZ* is not de-repressed along with Rh5 and remains restricted to pR8 (**d**). **e, f,** Expression of nuclear yR8 marker *wts-nlacZ* does not change in *rh6* mutants. In wild-type flies it is expressed with Rh6 (red) in yR8 and never in Rh5-expressing pR8 cells (blue) (**e**). In *rh6* mutants, *wts-nlacZ* remains in yR8 of 4-week-old flies as Rh5 is de-repressed (**f**).

switch involving mutual transcriptional repression between *warts* (*wts*) and *melted* (*melt*) (Supplementary Fig. 1b). During pupal development, Wts represses *melt* to specify yR8 photoreceptors. In response to an extrinsic signal originating in pR7, *melt* is upregulated in pR8, leading to repression of *wts* transcription and expression of Rh5 (ref. 8). Thus, Melt marks pR8 and Wts marks yR8 cells (Fig. 3c, e). In old *rh6* mutant flies, a *melt* reporter (*melt-nlacZ*) remains restricted to a subset of R8 cells, whereas Rh5 expression expands broadly to cells that do not express *melt-nlacZ* (Fig. 3d and Supplementary Fig. 9d). In addition, we do not observe downregulation of a *wts* reporter (*wts-nlacZ*) in yR8s of old *rh6* mutants, leading to co-expression of *wts* with ectopic Rh5 (Fig. 3f and Supplementary Fig. 9d). Although maintenance of *rh6-lacZ* and *wts-nlacZ* could potentially be due to perdurance of β Gal protein, lack of de-repression of *melt-lacZ* argues that loss of *rh6* function does not affect the identity of yR8 in old flies. Moreover, it shows that *melt* is not involved in Rh5 de-repression. Thus, in *rh6* mutants, the yR8 fate is specified normally and remains stable. This indicates that yR8 produces positive transcriptional regulatory inputs to which the *rh5* promoter can respond and which must be actively

repressed by the presence of Rh6. In contrast to the way pR8 *rh5*-expressing photoreceptor fate is established, these inputs do not depend on extrinsic signals from R7 cells because, as described earlier, the absence of R7s in *sev* mutants does not suppress the *rh6* mutant phenotype.

yR8 cells are not the only photoreceptors expressing Rh6. The larval eye, Bolwig's organ, is composed of about 12 photoreceptors^{21,22}. Four primary photoreceptors express Rh5 whereas the eight secondary photoreceptors express Rh6 (Supplementary Fig. 6a). During metamorphosis, secondary photoreceptors die while the primary photoreceptors downregulate Rh5 and upregulate Rh6 (ref. 23). The newly Rh6-expressing cells form the eyelet, an adult extra-retinal visual organ^{24,25} (Supplementary Fig. 6c). In *rh6* mutants, neither the secondary Bolwig photoreceptors nor the eyelet photoreceptors ever express Rh5 and are thus devoid of any rhodopsin (Supplementary Fig. 6b, d). Therefore, in contrast to the retina, Rh6 is not necessary for exclusion of Rh5 expression in the eyelet, consistent with the view that expression of Rh5 and Rh6 is under distinct control mechanisms in the Bolwig's organ/eyelet and in the adult retina²². This result, together with the absence of Rh5 de-repression in R7s of *rh3* and *rh4* mutants, argues that, in the absence of a rhodopsin signal, de-repression of Rh5 can only occur in yR8 photoreceptors.

Because Rh5 is only de-repressed in yR8s of *rh6* mutants, it is possible that the repressive signal is generated uniquely by Rh6. Therefore, we tested whether the *rh6* mutant phenotype in yR8s could be rescued by rhodopsins other than Rh6. We used *rh6-Gal4* to drive expression of UAS-Rh1, -Rh3, -Rh4 or -Rh6 in *rh6* mutants. In every case, the de-repression was rescued and little or no Rh5 expression was detectable in yR8 photoreceptors (Fig. 4a, b and Supplementary Figs 7b–e and 9e). Expression of UAS-Rh5, as with Rh1, Rh3 and Rh6, also largely blocked de-repression of the *rh5>GFP* reporter in *rh6* mutants (Fig. 4c and Supplementary Fig. 7f–i, 9f), suggesting that a generic *Drosophila* rhodopsin signal is sufficient to maintain exclusion of Rh5 in yR8 cells. Because these transgenes are controlled by the *rh6* promoter, they are expressed only after specification of the yR8 subtype, further arguing that the signal is only required for the maintenance of the exclusion of Rh5, and not for yR8 subtype specification. In addition, negative regulation by Rh5 of its own expression in yR8 could provide an explanation for why the levels of Rh5 expression in yR8 of *rh6* mutants are generally lower than in wild-type pR8 cells.

The requirement for a rhodopsin-dependent signal to maintain repression of *rh5* in yR8s led us to ask whether activation of Rh6 by light is involved in this process. We maintained wild-type flies in complete darkness for more than 2 weeks starting at mid-pupal stages. In these flies, a significant proportion (~12%, Supplementary Table 2) of the Rh6-expressing yR8s also express low levels of Rh5 (Fig. 4d, e and Supplementary Fig. 9g), which is not observed in old wild-type flies reared in the light. Interestingly, this de-repression of Rh5 occurs predominantly in the dorsal retina (Supplementary Table 2 and Supplementary Fig. 9g), indicating an underlying spatial variation in Rh5 de-repression. In contrast, Rh6 is not de-repressed in pR8s of dark-reared flies. Thus, it appears that adult yR8 photoreceptor neurons remain plastic with respect to rhodopsin exclusion and that simply preventing activation of Rh6 by light can evoke Rh5 expression in yR8s. This de-repression of Rh5, however, is substantially weaker than in *rh6* mutants. This could indicate that either activated Rh6 somehow accumulates in the dark and is able to partly repress *rh5* or that Rh6 retains a residual ability to repress *rh5* without being activated by light. These alternatives are consistent with the observation that partial reduction of Rh6 protein through RNA interference can lead to de-repression of *rh5* (Supplementary Fig. 5). Hence, *rh5* repression is sensitive to the level/activity of Rh6.

The role of light and interchangeability of rhodopsins in controlling expression of *rh5* raised the possibility that components of the phototransduction cascade (reviewed in ref. 5) might play a role in repression of *rh5*. In flies, activated rhodopsin converts the $G_{\alpha q}$ subunit of a

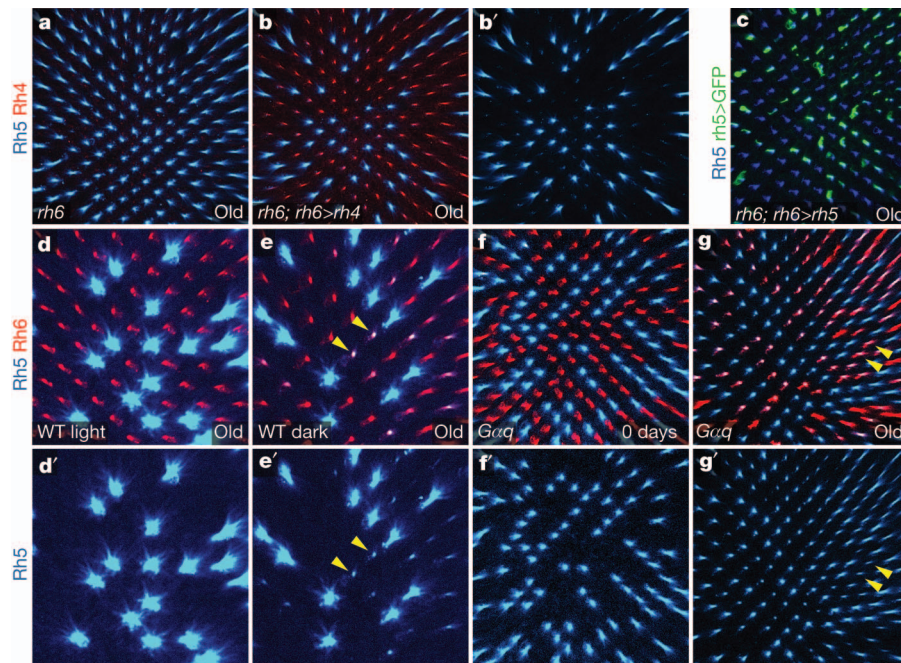


Figure 4 | Part of the phototransduction pathway is required to maintain repression of Rh5. **a, b,** Forced expression of Rh4 (red) in yR8 with *rh6*-Gal4 in *rh6*¹ mutants prevents Rh5 (blue) de-repression (**b**) observed in *rh6*¹ mutant flies (**a**). **c,** Forced expression of Rh5 (blue) in yR8 with *rh6*-Gal4 in *rh6*¹ mutants prevents *rh5*>GFP (green) de-repression observed in *rh6*¹ mutant flies (compare with Fig. 2f and Supplementary Fig. 7f). **d, e,** Dark-reared flies partly de-repress Rh5 in yR8 photoreceptors. In the light, wild-type flies do not

de-repress Rh5 (blue) in Rh6-expressing yR8s (red) (**d**). After 2–3 weeks in complete darkness, a significant number of yR8 cells of wild-type flies express low levels of Rh5 in addition to Rh6 (arrowheads, **e**). **d, e,** Close-ups of dorsal retinas, just dorsal to the equator. **f, g,** *Gαq* is required to maintain repression of Rh5 in yR8. In 2- to 3-week-old (**g**), but not in just eclosed (**f**), *Gαq*¹ mutants, Rh5 (blue) is expressed in yR8 and thus is co-expressed (arrowheads) with Rh6 (red). **b', d'–g',** Rh5 expression alone as in **b, d–g**.

heterotrimeric G-protein to a GTP-bound form which dissociates from the $G_{\beta\gamma}$ dimer and activates phospholipase C encoded by the *norpA* gene. Phospholipase C then catalyses hydrolysis of PIP₂, which leads to the activation of TRPC channels²⁶, inflow of Ca²⁺ and depolarization of the photoreceptors. We asked whether components of this phototransduction cascade mediate the *rh5*-repressive signal. In *Gαq*¹ hypomorphic mutants, Rh5 is expressed normally in young flies but becomes de-repressed in yR8 as the flies age (Fig. 4f, g and Supplementary Fig. 9h), a phenotype similar to that of *rh6* mutants. This results in the co-expression of Rh5 and Rh6 in yR8 cells. However, removal of neither phospholipase C (in *norpA*³⁶ mutants) nor TRPC channels (in *trpl*³⁰²; *trp*³⁰¹ double mutants) leads to de-repression of Rh5 in yR8s of old flies (Supplementary Fig. 8, 9h). The observation that *Gαq*, but not the rest of the phototransduction cascade, is important for the *rh5*-repressive signal indicates a bifurcation of the phototransduction and *rh5*-repression pathways downstream of *Gαq*. Alternatively, *Gαq* might function genetically upstream of Rh6, for example, by stabilizing the Rh6 protein. In either case, Rh6 uses a pathway distinct from phototransduction to repress *rh5*. Importantly, the *Gαq* mutant phenotype and de-repression of Rh5 in dark-raised wild-type flies further support the idea that maintenance of *rh5* repression requires the activity of the Rh6 protein.

Rhodopsins canonically act as sensory receptor proteins. However, Rh1 also has non-visual functions; it is required for the proper formation and maintenance of the rhabdomeres of R1–R6 photoreceptors^{27,28} and has recently been shown to be involved in thermotactic discrimination²⁹. We showed here a new and surprising role for Rh6: it represses transcription of an alternative receptor gene, *rh5*, and thereby maintains the sensory specificity of yR8. This mechanism prevents Rh5/Rh6 co-expression, which would broaden the sensitivity spectrum of yR8 photoreceptors³⁰, limiting the ability of the visual system to discriminate colours. Furthermore, change in the yR8 spectrum could lead to sensory confusion if the downstream neuronal circuits misinterpret the information they receive. The repression of

rh5 by Rh6 also illustrates a so far unappreciated plasticity of yR8 photoreceptor neurons, as revealed by de-repression of Rh5 in wild-type flies reared in darkness. Constant darkness could mimic special environmental conditions, natural for the fly, under which lowered Rh6 activity evokes expression of Rh5 in yR8 photoreceptors to change spectral properties of the eye, or simply to boost the overall light response. Finally, the fact that we found two different *rh6* mutations in laboratory stocks raises a possibility that mutations in the *rh6* gene are also frequent in the natural population.

Repression of *rh5* by Rh6 is reminiscent of the control of olfactory receptor genes in vertebrate olfactory neurons¹⁴, which encode G-protein-coupled receptors similar to Rhodopsins. With rare exception, each olfactory neuron expresses only one allele of one olfactory receptor gene. This exclusion mechanism is not well understood, but requires an active olfactory receptor to generate a feedback signal for repression of other olfactory receptor genes^{10–14}. There, however, the feedback control of exclusion appears to be a common mechanism in all olfactory neurons, in contrast to the fly retina where only *rh5* is regulated by another rhodopsin, and only in the yR8 photoreceptor subtype.

Our findings show that cross-repression of sensory receptors is not unique to vertebrate chemosensory systems, but could be a more widely implemented mechanism by which mature sensory neurons, or other G-protein-coupled receptor-expressing cells, maintain their functional specificity. The relative simplicity of yR8 photoreceptors as a system should allow us to uncover the molecular details by which a G-protein-coupled receptor can exclude expression of other seven-transmembrane receptors in the same cell.

METHODS SUMMARY

Flies were raised on standard corn meal/molasses/agar medium at room temperature (24 °C) in ambient laboratory light except for RNAi experiments (at 29 °C) and dark isolation experiments (in complete darkness). Dissected adult retinas were stained wholemount with specific primary antibodies and then with Alexa Fluor-conjugated secondary antibodies (Molecular Probes). Larval eyes were stained as in ref. 22. *In situ* hybridization on cryo-sectioned adult retinas was

performed with DIG-labelled antisense probe transcribed from *rh5* 3' untranslated region (UTR) as described in ref. 7. Samples were imaged using Leica TCS SP2 and SP5 confocal microscopes. Images were processed and counts performed using Leica Confocal Software, Adobe Photoshop and Fiji software. For real-time PCR, RNA was purified from 20 flies per sample and complementary DNA (cDNA) amplified using SYBR-Green PCR Mix (Stratagene).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 October 2010; accepted 15 August 2011.

Published online 9 October 2011.

- Mombaerts, P. Axonal wiring in the mouse olfactory system. *Annu. Rev. Cell Dev. Biol.* **22**, 713–737 (2006).
- Komiyama, T. & Luo, L. Development of wiring specificity in the olfactory system. *Curr. Opin. Neurobiol.* **16**, 67–73 (2006).
- Morey, M. *et al.* Coordinate control of synaptic-layer specificity and rhodopsins in photoreceptor neurons. *Nature* **456**, 795–799 (2008).
- Rister, J. & Desplan, C. The retinal mosaics of opsin expression in invertebrates and vertebrates. *Dev. Neurobiol.* doi:10.1002/dneu.20905 (2011).
- Wang, T. & Montell, C. Phototransduction and retinal degeneration in *Drosophila*. *Pflügers Arch.* **454**, 821–847 (2007).
- Franceschini, N., Kirschfeld, K. & Minke, B. Fluorescence of photoreceptor cells observed *in vivo*. *Science* **213**, 1264–1267 (1981).
- Mazzoni, E. O. *et al.* Iroquois complex genes induce co-expression of rhodopsins in *Drosophila*. *PLoS Biol.* **6**, e97 (2008).
- Mikeladze-Dvali, T. *et al.* The growth regulators warts/lats and melted interact in a bistable loop to specify opposite fates in *Drosophila* R8 photoreceptors. *Cell* **122**, 775–787 (2005).
- Wernet, M. F. *et al.* Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature* **440**, 174–180 (2006).
- Serizawa, S. *et al.* Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science* **302**, 2088–2094 (2003).
- Shykind, B. M. *et al.* Gene switching and the stability of odorant receptor gene choice. *Cell* **117**, 801–815 (2004).
- Lewcock, J. W. & Reed, R. A feedback mechanism regulates monoallelic odorant receptor expression. *Proc. Natl Acad. Sci. USA* **101**, 1069–1074 (2004).
- Feinstein, P., Bozza, T., Rodriguez, I., Vassalli, A. & Mombaerts, P. Axon guidance of mouse olfactory sensory neurons by odorant receptors and the $\beta 2$ adrenergic receptor. *Cell* **117**, 833–846 (2004).
- Fuss, S. H. & Ray, A. Mechanisms of odorant receptor gene choice in *Drosophila* and vertebrates. *Mol. Cell. Neurosci.* **41**, 101–112 (2009).
- Cook, T., Pichaud, F., Sonnevill, R., Papatsenko, D. & Desplan, C. Distinction between color photoreceptor cell fates is controlled by Prospero in *Drosophila*. *Dev. Cell* **4**, 853–864 (2003).
- Harris, W. A., Stark, W. S. & Walker, J. A. Genetic dissection of the photoreceptor system in the compound eye of *Drosophila melanogaster*. *J. Physiol. (Lond.)* **256**, 415–439 (1976).
- Chou, W. H. *et al.* Patterning of the R7 and R8 photoreceptor cells of *Drosophila*: evidence for induced and default cell-fate specification. *Development* **126**, 607–616 (1999).
- Papatsenko, D., Sheng, G. & Desplan, C. A new rhodopsin in R8 photoreceptors of *Drosophila*: evidence for coordinate expression with Rh3 in R7 cells. *Development* **124**, 1665–1673 (1997).
- Chou, W. H. *et al.* Identification of a novel *Drosophila* opsin reveals specific patterning of the R7 and R8 photoreceptor cells. *Neuron* **17**, 1101–1115 (1996).
- Tahayato, A. *et al.* Otd/Crx, a dual regulator for the specification of ommatidia subtypes in the *Drosophila* retina. *Dev. Cell* **5**, 391–402 (2003).
- Green, P., Hartenstein, A. Y. & Hartenstein, V. The embryonic development of the *Drosophila* visual system. *Cell Tissue Res.* **273**, 583–598 (1993).
- Sprecher, S. G., Pichaud, F. & Desplan, C. Adult and larval photoreceptors use different mechanisms to specify the same Rhodopsin fates. *Genes Dev.* **21**, 2182–2195 (2007).
- Sprecher, S. G. & Desplan, C. Switch of rhodopsin expression in terminally differentiated *Drosophila* sensory neurons. *Nature* **454**, 533–537 (2008).
- Hofbauer, A. & Buchner, E. Does *Drosophila* have seven eyes? *Naturwissenschaften* **76**, 335–336 (1989).
- Yasuyama, K. & Meinertzhagen, I. A. Extraretinal photoreceptors at the compound eye's posterior margin in *Drosophila melanogaster*. *J. Comp. Neurol.* **412**, 193–202 (1999).
- Huang, J. *et al.* Activation of TRP channels by protons and phosphoinositide depletion in *Drosophila* photoreceptors. *Curr. Biol.* **20**, 189–197 (2010).
- O'Tousa, J. E., Leonard, D. S. & Pak, W. L. Morphological defects in oraJK84 photoreceptors caused by mutation in R1–6 opsin gene of *Drosophila*. *J. Neurogenet.* **6**, 41–52 (1989).
- Kumar, J. P. & Ready, D. F. Rhodopsin plays an essential structural role in *Drosophila* photoreceptor development. *Development* **121**, 4359–4370 (1995).
- Shen, W. L. *et al.* Function of rhodopsin in temperature discrimination in *Drosophila*. *Science* **331**, 1333–1336 (2011).
- Arikawa, K., Mizuno, S., Kinoshita, M. & Stavenga, D. G. Coexpression of two visual pigments in a photoreceptor causes an abnormally broad spectral sensitivity in the eye of the butterfly *Papilio xuthus*. *J. Neurosci.* **23**, 4527–4532 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Blau, B. Collins, M. Cols, T. Erclik, S.H. Fuss, D. Jukam, J.P. Kumar, E. Laufer, H.-S. Li, B. Minke, C. Montell, F. Pichaud, J. Rister and A. Tomlinson for suggestions and comments on the manuscript, V. Douard for help with qRT-PCR, J. Goodness for help identifying *rh6^{ts}* allele, S.G. Britt, P.J. Dolph, P.R. Hiesinger, F. Pichaud, N. Pinal, D.F. Ready, C.S. Zuker, and the Bloomington *Drosophila* Stock Center for flies or antibodies. This work was funded by the National Institutes of Health R01EY13012 to C.D. and F32EY016309 to D.V.

Author Contributions D.V., E.O.M. and C.D. conceived the experiments; D.V. and E.O.M. designed and performed experiments in adult flies; S.G.S. designed and performed experiments in larvae; K.B. performed RNAi experiments; R.J.J., P.L., N.V. and A.C. contributed reagents; D.V. and C.D. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.D. (cd38@nyu.edu).

METHODS

Flies were raised on standard cornmeal/molasses/agar medium at room temperature (24 °C) in ambient laboratory light unless otherwise noted. RNAi experiments were performed at 29 °C. For dark isolation experiments, flies were reared in a lightproof box, and for ageing transferred between vials in complete darkness starting at mid-pupal stages (before rhodopsin expression³¹).

Drosophila strains. For wild-type controls we used $y^1 w^{67}$; *Sp/CyO*; *wt^{isoB}* flies, where 'isoB' represents an isogenized wild-type third chromosome.

We used the following *rh6* alleles.

The *rh6¹* allele¹⁵ is found in many commonly used laboratory fly strains. The existence of this mutation in common stocks was originally pointed out to us by S. Britt. This mutation is present on some TM6B balancer chromosomes and in the reference fly strain sequenced for the published fly genome¹⁵ (BDGP release 5.29). The mutation replaces 21 bases (lower case in TGACCATCATCTCTctac tggcacatcatgaaggTATGACATTCGTTA) at the end of the first exon with two As, removing a splice donor site and introducing a stop codon immediately afterwards. This results in the truncation of the open reading frame within the fifth transmembrane domain of the presumptive protein. The original allele was backcrossed into *wt^{isoB}* background (see above) four times.

We identified *rh6⁵* as a mutation in a stock from the Bloomington *Drosophila* Stock Center (Stock 1385, named genotype $z^{77h} w^{67c23}$) which mapped to the third chromosome. Sequencing of *rh6* locus revealed a 58-bp deletion upstream of the *rh6* transcription start site, which removes sequence AGCGGCAATCGAAAGCC CAATTCGAACGGTTAGCTTTGGATTGGCCAAGTGCCGGCTA within the *rh6* promoter. We named this mutation after the singer Frank Sinatra, for his nickname "Ol' blue eyes", because eyes of old *rh6⁵* mutant flies broadly express the blue-sensitive rhodopsin, Rh5.

The deficiency used in this study that covers *rh6* gene, *Df(3R)Exel6174*, was generated by Exelixis and spans 3R:11154443–11154444 ... 11363188 (ref. 32).

To generate flies with a *rh6* genomic rescue fragment, *C(rh6⁺)*, the *rh6* sequence was PCR-amplified from genomic DNA of $y^1 w^{67}$; *Sp/CyO*; *wt^{isoB}* flies with dv173 (ACAAGCTTACCTACAAGAGCACCAGTCC) and dv174 (ACGAATTCA CCTCGGCTGAACACCTAC) primers to produce a 2,575-bp genomic fragment (ACCTACAAGAGCACCAGTCC ... GTAGGTGTTGAGCGCGAGGT) with HindIII and EcoRI flanking sites. PCR product was ligated into HindIII–EcoRI sites of pBS-loxP-w-lox2272 vector³³. Cre-recombinase-mediated integration was used to insert this construct into lox landing site A11 (on the second chromosome, S. Small, personal communication). A single integration occurred without replacement of y^+ marker of the landing site. Successful transformation was confirmed with antibody stain for Rh6 protein in wholemount retinas: normal Rh6 expression was detected in *rh6* mutant background.

UAS-rh6RNAi (transformant 102152) was obtained from the Vienna *Drosophila* RNAi Center³⁴.

We generated the following additional mutants.

rh3¹ mutant (a nucleotide change C278T resulting in Q46* truncation) was obtained by TILLING (Seattle TILLING Project)³⁵. The mutation was backcrossed into wild-type background four times (confirmed by genomic PCR and by stain of wholemount retinas with anti-Rh3 antibody).

rh4¹ mutant (a nucleotide change T727A resulting in Y203* truncation between fourth and fifth transmembrane domains) was obtained by TILLING (Seattle TILLING Project)³⁵. The mutation was back crossed into wild-type background four times (confirmed by genomic PCR and by stain of wholemount retinas with anti-Rh4 antibody).

We generated the following transgenic lines.

rh5>GFP flies carry two transgenes recombined on the second chromosome: *rh5-lexA* and *lexAop-GFP*.

rh5-lexA. *lexA* (from pBS-lexA SV40 3' UTR³⁶) was cloned into pBS-LoxP-white-Lox2272 (ref. 33) and named LexA-Lox, a 740-bp fragment of *rh5* promoter which ends 23 bases upstream of ATG (TCGGAATATGTCGTGCAAGTGTTC ... AATGTGCACTGCAAAGGAACTA; fly genome: 12007686 ... 12008425) was PCR amplified from genomic DNA using oBJ109 (TCGGAATATGT CGTGCAGTG) and oBJ140 (TAGTTTCCTTTGCGAGGTCGAC) and cloned into the PCRII-TOPO vector (Invitrogen). The *rh5* promoter was cut with ClaI, blunted and subcloned into the LexA-Lox, which was cut with SpeI and blunted. Cre-recombinase-mediated cassette exchange was used to insert this construct onto the second chromosome³³.

lexAop-GFP. GFP with SV40 3' UTR was PCR amplified from the pIRES2-eGFP vector (Clontech) with the primers oBJ78 (TAATACTAGTATGGTGA GCAAGGGCGAGGAG) and oBJ79 (GTCAGGATCCACCACAAGTAGAATG CAGTG) and cloned into the PCRII-TOPO vector (Invitrogen). The GFP-SV40 3' UTR was subcloned into the pLOT vector (containing *lexAop*)³⁶ using the EcoRI site.

UAS-Rh1. EcoRI–KpnI fragment of *rh1* cDNA (containing sequence spanned by GGCAGGTTTCCAACGACCAATCGC ... AAGGACAAAAAACTCA AC + 15A) from rh1-pFLC-1 plasmid (*Drosophila* Genomics Resource Center clone RH01460 (ref. 37)) was ligated into EcoRI–KpnI sites of pUASTattB vector³⁸ to produce pDV131 plasmid. ϕ C31-mediated integration was used to insert this construct into second chromosome landing sites *attP-51D*, *attP-58A* and *attP40* (refs 38, 39). w^+ and *3xP3-RFP* markers of *attP-51D* and *attP-58A* landing sites were removed through lox-mediated recombination by crossing in Cre recombinase transgene³⁸.

UAS-Rh3. EcoRI–XhoI fragment of *rh3* cDNA (containing sequence spanned by CAGACCGGAGCATGGAGTCCGGTA ... AATATAGTAAATACAGC AAGCT + 19A) from rh3-pOT2 (*Drosophila* Genomics Resource Center clone GH02505 (ref. 37)) was ligated into EcoRI–XhoI sites of pUASTattB vector³⁸ to produce pDV133 plasmid. ϕ C31-mediated integration was used to insert this construct into second chromosome landing sites *attP-51D*, *attP-58A* and *attP40* (refs 38, 39). w^+ and *3xP3-RFP* markers of *attP-51D* and *attP-58A* landing sites were removed through lox-mediated recombination by crossing in Cre recombinase transgene³⁸.

UAS-Rh4. Cloned EcoRI–KpnI fragment of *rh4* cDNA from rh4-pFLC-1 (*Drosophila* Genomics Resource Center clone RH33063 (ref. 37)) was ligated into pUASTattB vector³⁸. To correct a frameshift in the sequence, EcoRI–BglII fragment was replaced with cDNA fragment that had a longer 5' UTR. The resulting pDV134 plasmid contained *rh4* cDNA sequence spanned by CAGAGCGAAAC GGGTAGCGGT ... AACCTATTGCAAACGAAGTAG + 16A. ϕ C31-mediated integration was used to insert this construct into second chromosome landing sites *attP-51D* and *attP40* (refs 38, 39). w^+ and *3xP3-RFP* markers of *attP-51D* landing site were removed through lox-mediated recombination by crossing in Cre recombinase transgene³⁸.

UAS-Rh5. EcoRI–XhoI fragment of *rh5* cDNA (containing sequence spanned by CGGAGGCCAGAATGTGACCT ... TACAAACCAAAAAAGTTGGCA TT + 78A) from rh5-pOT2 (*Drosophila* Genomics Resource Center clone GH28578 (ref. 37)) was ligated into EcoRI–XhoI sites of pUASTattB vector³⁸ to produce pDV135 plasmid. ϕ C31-mediated integration was used to insert this construct into second chromosome landing site *attP40* (ref. 39).

UAS-Rh6. It has proved difficult to generate a *UAS-Rh6* cDNA construct expressing high levels of Rh6. Therefore, we cloned a PCR-amplified genomic (with introns) fragment of *rh6* gene downstream of transcriptional start site (containing sequence spanned by CAGGATTGCGCGGAGTTCGCGT ... ACAG CAATTGATACAAAATC) into EcoRI–KpnI sites of pUASTattB vector³⁸ to produce pDV160 plasmid. ϕ C31-mediated integration was used to insert this construct into second chromosome landing site *attP40* (ref. 39).

We used the following other strains.

G_{284}^1 (ref. 40), *norpA³⁶* (ref. 41), *rh5²* (ref. 42), *sev¹⁴* (ref. 43), *trp³⁰²*; *trp³⁰¹* (ref. 44), *melt-nlacZ⁸*, *rh6-Gal4* (ref. 20), *rh6-lacZ⁴⁵* and *wt5-nlacZ^{46,47}*.

Antibodies. Antibodies and dilutions used were as follows: mouse anti-Rh1 (1:10) (DSHB, clone 4C5); mouse anti-Rh3 (1:10) and mouse anti-Rh5 (1:100) (gifts from S. Britt); rabbit anti-Rh4 (1:100) (gift from C. Zuker); rabbit anti-Rh6 (1:2,000) (ref. 20); goat anti- β Gal (1:5,000) (Biogenesis); mouse anti- β Gal (1:500) (Promega); rat anti-Elav (1:40) (DSHB, clone Rat-Elav-7E8A10); sheep anti-GFP (1:500) (AbD Serotec); rabbit anti-GFP (1:800) (Biogenesis). Secondary antibodies raised in donkey and goat were Alexa-Fluor-conjugated (Alexa Fluor 488 at 1:1,000, Alexa Fluor 555 at 1:750, Alexa Fluor 647 at 1:500) (Molecular Probes). Alexa Fluor 488-conjugated phalloidin was used to visualize rhabdomeres (1:100, Molecular Probes).

Stains. Adult retinas were dissected out in phosphate buffered saline (PBS), fixed for 15 min with 4% formaldehyde at room temperature, washed three times in PBS, and incubated with the primary antibodies diluted in Block (PBS, 0.1% Triton-X-100, 2% horse serum) overnight at 4 °C. After two rinses and two 1-h washes with PBT (PBS, 0.3% Triton-X-100), the retinas were incubated overnight at 4 °C with secondary antibodies diluted in Block. Retinas were rinsed twice and after two 1-h washes with PBT, were mounted in SlowFade Gold (Invitrogen). Antibody staining for larval eye was performed as described in ref. 22. *In situ* hybridization for cryo-sectioned adult retinas was performed as described in ref. 7 with DIG-labelled antisense probe transcribed from cloned *rh5* 3' UTR region (bp 900–1411). Samples were imaged using Leica TCS SP2 and SP5 confocal microscopes. Images were processed using Leica Confocal Software (LCS), Adobe Photoshop and Fiji software.

Counting. Optical sections were photographed approximately 10 μ m distal to R8 nuclei in the centre of the retina. The portion of the image of the retina section containing R8 rhabdomeres was defined as area populated with Rh5-positive cells. The number of Rh5-expressing R8s and the total number of R8s (represented by ommatidia visualized with phalloidin) in this area were counted using Fiji software with Cell Counter plug in.

RNA analysis. RNA was purified from each sample of about 20 flies with TRIzol (Invitrogen), RNeasy mini columns (Qiagen) and treated with DNaseI (Qiagen). Three micrograms of total RNA was reverse transcribed with oligo(dT)20 and SuperScript III Reverse Transcriptase (Invitrogen). The cDNA was amplified in duplicate reactions using SYBR-Green PCR Mix (Stratagene) by real-time PCR. Primers used are listed in Supplementary Table 3. Target gene levels were normalized to levels of *rp49* mRNA⁴⁸ and expressed relative to levels in 0-day-old wild-type flies. At least three independent replicates were averaged for each experimental condition.

31. Earl, J. B. & Britt, S. G. Expression of *Drosophila* rhodopsins during photoreceptor cell differentiation: insights into R7 and R8 cell subtype commitment. *Gene Expr. Patterns* **6**, 687–694 (2006).
32. Parks, A. L. *et al.* Systematic generation of high-resolution deletion coverage of the *Drosophila melanogaster* genome. *Nature Genet.* **36**, 288–292 (2004).
33. Oberstein, A., Pare, A., Kaplan, L. & Small, S. Site-specific transgenesis by Cre-mediated recombination in *Drosophila*. *Nature Methods* **2**, 583–585 (2005).
34. Dietzl, G. *et al.* A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* **448**, 151–156 (2007).
35. Till, B. J. *et al.* High-throughput TILLING for functional genomics. *Methods Mol. Biol.* **236**, 205–220 (2003).
36. Lai, S. L. & Lee, T. Genetic mosaic with dual binary transcriptional systems in *Drosophila*. *Nature Neurosci.* **9**, 703–709 (2006).
37. Stapleton, M. *et al.* The *Drosophila* gene collection: identification of putative full-length cDNAs for 70% of *D. melanogaster* genes. *Genome Res.* **12**, 1294–1300 (2002).
38. Bischof, J., Maeda, R. K., Hediger, M., Karch, F. & Basler, K. An optimized transgenesis system for *Drosophila* using germ-line-specific ϕ C31 integrases. *Proc. Natl Acad. Sci. USA* **104**, 3312–3317 (2007).
39. Markstein, M., Pitsouli, C., Villalta, C., Celniker, S. E. & Perrimon, N. Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nature Genet.* **40**, 476–483 (2008).
40. Scott, K., Becker, A., Sun, Y., Hardy, R. & Zuker, C. Gq α protein function in vivo: genetic dissection of its role in photoreceptor cell physiology. *Neuron* **15**, 919–927 (1995).
41. Bloomquist, B. T. *et al.* Isolation of a putative phospholipase C gene of *Drosophila*, *norpA*, and its role in phototransduction. *Cell* **54**, 723–733 (1988).
42. Yamaguchi, S., Wolf, R., Desplan, C. & Heisenberg, M. Motion vision is independent of color in *Drosophila*. *Proc. Natl Acad. Sci. USA* **105**, 4910–4915 (2008).
43. Gerresheim, F. Isolation of *Drosophila melanogaster* mutants with a wavelength-specific alteration in their phototactic response. *Behav. Genet.* **18**, 227–246 (1988).
44. Niemeyer, B. A., Suzuki, E., Scott, K., Jalink, K. & Zuker, C. S. The *Drosophila* light-activated conductance is composed of the two channels TRP and TRPL. *Cell* **85**, 651–659 (1996).
45. Papatsenko, D., Nazina, A. & Desplan, C. A conserved regulatory element present in all *Drosophila* rhodopsin genes mediates Pax6 functions and participates in the fine-tuning of cell-specific expression. *Mech. Dev.* **101**, 143–153 (2001).
46. Justice, R. W., Zilian, O., Woods, D. F., Noll, M. & Bryant, P. J. The *Drosophila* tumor suppressor gene *warts* encodes a homolog of human myotonic dystrophy kinase and is required for the control of cell shape and proliferation. *Genes Dev.* **9**, 534–546 (1995).
47. Xu, T., Wang, W., Zhang, S., Stewart, R. A. & Yu, W. Identifying tumor suppressors in genetic mosaics: the *Drosophila* *lats* gene encodes a putative protein kinase. *Development* **121**, 1053–1063 (1995).
48. O'Keefe, L. V. *et al.* *Drosophila* orthologue of WWOX, the chromosomal fragile site FRA16D tumour suppressor gene, functions in aerobic metabolism and regulates reactive oxygen species. *Hum. Mol. Genet.* **20**, 497–509 (2011).

Verbal and non-verbal intelligence changes in the teenage brain

Sue Ramsden¹, Fiona M. Richardson¹, Goulven Josse¹, Michael S. C. Thomas², Caroline Ellis¹, Clare Shakeshaft¹, Mohamed L. Seghier¹ & Cathy J. Price¹

Intelligence quotient (IQ) is a standardized measure of human intellectual capacity that takes into account a wide range of cognitive skills¹. IQ is generally considered to be stable across the lifespan, with scores at one time point used to predict educational achievement and employment prospects in later years¹. Neuroimaging allows us to test whether unexpected longitudinal fluctuations in measured IQ are related to brain development. Here we show that verbal and non-verbal IQ can rise or fall in the teenage years, with these changes in performance validated by their close correlation with changes in local brain structure. A combination of structural and functional imaging showed that verbal IQ changed with grey matter in a region that was activated by speech, whereas non-verbal IQ changed with grey matter in a region that was activated by finger movements. By using longitudinal assessments of the same individuals, we obviated the many sources of variation in brain structure that confound cross-sectional studies. This allowed us to dissociate neural markers for the two types of IQ and to show that general verbal and non-verbal abilities are closely linked to the sensorimotor skills involved in learning. More generally, our results emphasize the possibility that an individual's intellectual capacity relative to their peers can decrease or increase in the teenage years. This would be encouraging to those whose intellectual potential may improve, and would be a warning that early achievers may not maintain their potential.

An individual's abilities and capacity to learn can be partly captured by the use of verbal and non-verbal (henceforth performance) intelligence tests. IQ provides a standardized method for measuring intellectual abilities and is widely used within education, employment and clinical practice. In the absence of neurological insult or degenerative conditions, IQ is usually expected to be stable across lifespan, as evidenced by the fact that IQ measurements made at different points in an individual's life tend to correlate well^{1,2}. Nevertheless, strong correlations over time disguise considerable individual variation; for example, a correlation coefficient of 0.7 (which is not unusual with verbal IQ) still leaves over 50% of the variation unexplained. The study that we report here tested whether variation in a teenager's IQ over time correlated with changes in brain structure. If it did, this would provide construct validity for the increase or decrease of IQ in the teenage years, because if IQ changes correspond to structural brain changes then they are unlikely to represent measurement error in the IQ tests. In addition, if verbal and performance skills change at different rates in different individuals, the neural markers for verbal and performance IQ changes could in principle be dissociated. This would overcome two of the challenges faced by previous studies of between-subject variability in IQ measures at a given time point: verbal and performance IQ are tightly correlated in individuals, so it has been hard to identify neural structures corresponding to each^{3,4}; and there are many sources of between-subject variance in brain structure (for example gender, age, size and handedness) that hide the relevant differences.

Our participants were 33 healthy and neurologically normal adolescents with a deliberately wide and heterogeneous mix of abilities (see Supplementary Information for details and the implications of our sampling for the generalizability of our conclusions). They were first tested in 2004 ('time 1') when they were 12–16 yr old (mean, 14.1 yr). Testing was repeated in 2007/2008 ('time 2') when the same individuals were 15–20 yr old (mean, 17.7 yr). See Table 1 for further details of the participants. During the intervening years, there were no testing sessions, and participants and their parents had no knowledge that they would be invited back for further testing. On both test occasions, each participant had a structural brain scan using magnetic resonance imaging (MRI) and had their IQ measured using the Wechsler Intelligence Scale for Children (WISC-III) at time 1 and the Wechsler Adult Intelligence Scale (WAIS-III) at time 2 (see Supplementary Information for details). These two widely used, age-appropriate assessments⁵ produce strongly correlated results at a given time point, consistent with them measuring highly similar constructs⁶. Scores on individual subtests are standardized against age-specific norms and then grouped to produce separate measures of verbal IQ (VIQ) and performance IQ (PIQ), with VIQ encompassing those tests most related to verbal skills and PIQ being more independent of verbal skills. Nevertheless, VIQ and PIQ scores are very significantly correlated with each other across participants: in our sample, the correlations between VIQ and PIQ were $r = 0.51$ at time 1 and $r = 0.55$ at time 2 (in both cases, $n = 33$; $P < 0.01$). Full-scale IQ (FSIQ) is the composite of VIQ and PIQ and is regarded as the best measure of general intellectual capacity (the g factor) that has previously been shown to correlate with brain size and cortical thickness in a wide variety of frontal, parietal and temporal brain regions^{7,8}.

The wide range of abilities in our sample was confirmed as follows: FSIQ ranged from 77 to 135 at time 1 and from 87 to 143 at time 2, with averages of 112 and 113 at times 1 and 2, respectively, and a tight correlation across testing points ($r = 0.79$; $P < 0.001$). Our interest was in the considerable variation observed between testing points at the individual level, which ranged from -20 to $+23$ for VIQ, -18 to $+17$ for PIQ and -18 to $+21$ for FSIQ. Even if the extreme values of the published 90% confidence intervals are used on both occasions, 39% of the sample showed a clear change in VIQ, 21% in PIQ and 33% in FSIQ. In terms of the overall distribution, 21% of our sample showed

Table 1 | Participants' details

	Datum	Age	FSIQ	VIQ	PIQ
Time 1	Mean (s.d.)	14.1 (1.0)	112 (13.9)	113 (15.1)	108 (12.3)
	Min, max	12.6, 16.5	77, 135	84, 139	74, 137
Time 2	Mean (s.d.)	17.7 (1.0)	113 (14.0)	116 (18.0)	107 (9.6)
	Min, max	15.9, 20.2	87, 143	90, 150	83, 124
Correlation*	r	—	0.792†	0.809†	0.589†
Change	Mean (s.d.)	3.5 (0.2)	+1.0 (9.0)	+3.0 (10.6)	-1.0 (10.2)
	Min, max	3.3, 3.9	-18, +21	-20, +23	-18, +17

* Correlation coefficient between scores at times 1 and 2. † $P < 0.01$.

$n = 33$ (19 male, 14 female). s.d., standard deviation.

¹Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, UK. ²Developmental Neurocognition Laboratory, Department of Psychological Sciences, Birkbeck College, University of London, London WC1E 7HX, UK.

a shift of at least one population standard deviation (15) in the VIQ measure, and 18% in the PIQ measure. However, only one participant had a shift of this magnitude in both measures, and, for that participant, one measure showed an increase and the other a decrease. This pattern is reflected in the absence of a significant correlation between the change in VIQ and the change in PIQ. The independence of

changes in these two measures allows us to investigate the effect of each without confounding influences from the other.

To test whether the observed IQ changes were meaningfully reflected in brain structure, we correlated them with changes in local brain structure. This within-subject correlation obviates the many possible sources of between-subject variance and may have sensitized our analysis to neural markers of VIQ and PIQ that have not previously been revealed. Given the distributed nature of brain regions associated with between-subject differences in FSIQ^{7–9}, regions of interest were not used in this analysis, and the results of the whole-brain analysis were only considered to be significant at $P < 0.05$ after familywise error correction in either height (peak signal at a single voxel) or extent (number of voxels that were significant at $P < 0.001$).

Using regression analysis, we studied the brain changes associated with a change in VIQ, PIQ or FSIQ (see Methods Summary for details). The results (Fig. 1) showed that changes in VIQ were positively correlated with changes in grey matter density (and volume) in a region of the left motor cortex that is activated by the articulation of speech¹⁰. Conversely, changes in PIQ were positively correlated with grey matter density in the anterior cerebellum (lobule IV), which is associated with motor movements of the hand^{11,12}. *Post hoc* tests that correlated structural change with change in each of the nine VIQ and PIQ subtest scores that were common in the WISC and WAIS assessments found that the neural marker for VIQ indexed constructs that were shared by all VIQ measures and that the neural marker for PIQ indexed constructs that were common to three of the four PIQ measures (Table 2). This indicates that our VIQ and PIQ markers indexed skills that were not specific to individual subtests. There were no other grey or white matter effects that reached significance in a whole-brain structural analysis of VIQ, PIQ or FSIQ. See Supplementary Information for details of further *post hoc* tests.

Our findings that VIQ changes were related to a motor speech region and PIQ changes were related to a motor hand region are consistent with previous claims that cognitive intelligence is partly dependent on sensorimotor skills^{13–18}. Using functional imaging in the same 33 participants performing a range of sensory, motor and language tasks, we confirmed that the left motor speech region identified in the VIQ structural analysis was more activated by articulation tasks (naming, reading and saying “one, two, three”) than by semantic or perceptual tasks that required a finger press response (see Supplementary Information for details). In contrast, a region very close to the anterior cerebellum region identified in the PIQ structural analysis was more activated during tasks involving finger presses than during tasks involving articulation. Figure 2 shows these results at both the group level and the individual level. The locations of the grey

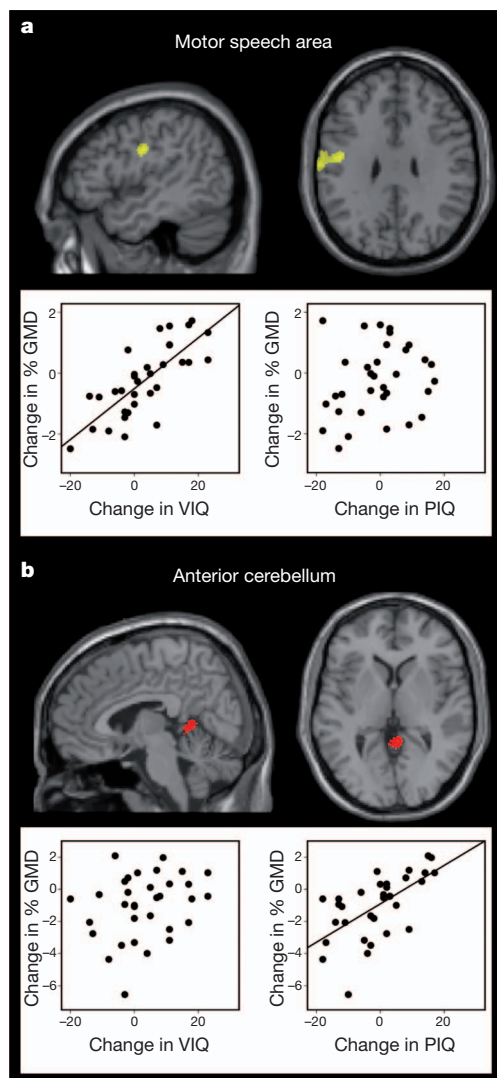


Figure 1 | Location of brain regions where grey matter changed with VIQ and PIQ. **a**, Correlation between change in grey matter density and change in VIQ (yellow) in the left motor speech region (peak in the left precentral gyrus at $x = -47$ mm, $y = -9$ mm, $z = +30$ mm, measured in Montreal Neurological Institute (MNI) space, with a Z-score of 5.2 and 681 voxels at $P < 0.001$). The corresponding effect on volume was slightly less significant (Z-score, 3.5; 118 voxels at $P < 0.001$). **b**, Correlation between change in PIQ (red) and change in grey matter density in the anterior cerebellum (peak at $x = +6$ mm, $y = -46$ mm, $z = +3$ mm, in MNI space, with a Z-score of 3.9 and 210 voxels at $P < 0.001$). Both effects were significant at $P < 0.05$ after familywise error correction for multiple comparisons in extent based on the number of voxels in a cluster that survived $P < 0.001$ uncorrected. In addition, the VIQ effect was significant at $P < 0.05$ after familywise error correction for multiple comparisons in height. The statistical threshold used in the figure ($P < 0.001$) illustrates the extent of the effects. Plots show the change in grey matter density versus the change in both VIQ and PIQ at the voxel with the highest Z-score in the appropriate region. Linear regression lines are shown for significant correlations. Changes in the motor speech region correlated with changes in VIQ but not changes in PIQ, whereas changes in the anterior cerebellum correlated with changes in PIQ but not changes in VIQ ($P < 0.001$). $n = 33$; GMD, grey matter density.

Table 2 | Correlations between grey matter density and score

Test type		Motor speech region (r)	Anterior cerebellum (r)
Verbal tests	Vocabulary	0.284*	0.142
	Similarities	0.438†	−0.021
	Arithmetic	0.477†	0.304‡
	Information	0.314‡	0.185
	Comprehension	0.541†	0.183
Non-verbal tests	Picture completion	0.038	0.363‡
	Digit symbol coding	0.003	−0.028
	Block design	0.000	0.306‡
	Picture arrangement	0.126	0.437†

*Trend (one-tailed) at $P = 0.0545$. †Significant (one-tailed) at $P < 0.01$. ‡Significant (one-tailed) at $P < 0.05$.

Correlations were calculated using changes in scaled (that is, age-adjusted) scores in the various subtests that were common to both the WISC and the WAIS. The change in grey matter density in the motor speech region correlated significantly with changes in scores in four of the five verbal subtests, and there was a near-significant trend in the fifth but it did not correlate significantly with changes in scores in any of the four tests that comprise PIQ. Conversely, the change in grey matter density in the anterior cerebellum correlated significantly with changes in scores in three of the four tests that comprise PIQ (the exception being the digit symbol coding test, which has a particular loading on processing speed) but correlated with changes in scores in only one of the verbal tests (the arithmetic test, which probably has the smallest verbal component of the verbal tasks).

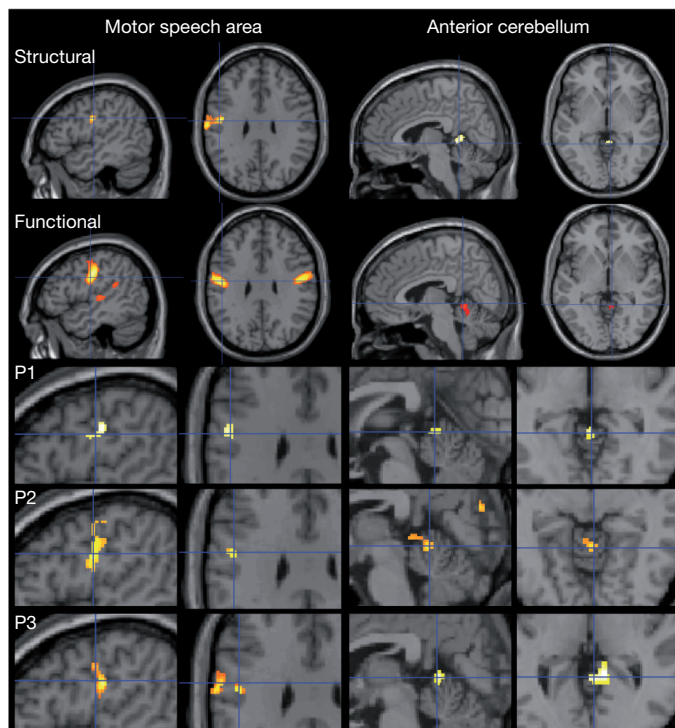


Figure 2 | Functional activations in the regions identified by the structural analysis. The motor speech region was more activated by articulation tasks than by finger press tasks ($x = -48$ mm, $y = -10$ mm, $z = +30$ mm (MNI); $t = 14.7$; $P < 0.05$ familywise-error-corrected for multiple comparisons across the whole brain), and corresponds to the region identified in the structural analysis for VIQ. These effects were consistently observed at the same coordinates for all individual subjects. In the three exemplar participants shown here (P1, P2, P3), the Z-scores were 3.9, 3.5 and 3.0, respectively. The anterior cerebellum region was more activated during finger presses than during articulation at both the group level (peak at $x = +6$ mm, $y = -48$ mm, $z = -4$ mm (MNI); Z-score, 3.7; 216 voxels at $P < 0.001$ corrected for multiple comparisons in extent) and the individual level (P1: $x = +12$ mm, $y = -48$ mm, $z = +2$ mm (MNI); Z-score, 3.7; P2: $x = +6$ mm, $y = -50$ mm, $z = -6$ mm (MNI); Z-score, 3.3; P3: $x = +12$ mm, $y = -46$ mm, $z = +2$ mm (MNI); Z-score, 4.9). In all cases, the activation peaks were identified from whole-brain analyses and the peak effects for the correlation with structure are illustrated with blue cross hairs in both the structural results and the functional results. This illustrates that the location of the structural effects is within the regions identified by the functional effects.

matter changes associated with VIQ and PIQ changes do not correspond to the anterior frontal and parietal regions associated with general intelligence⁷ (g factor). It may therefore be the case that g remains relatively constant across ages, but changes in the ability to perform individual subtests depend on changes in sensorimotor skills. It is also notable that although completion of the subtests comprising verbal and performance measures must implicate a network of brain regions, only structural changes in regions associated with sensorimotor skills showed correlations with changes in VIQ and PIQ.

The changes in brain structure that correlated with changes in IQ allow us to explain some of the variance in terms of brain development. Specifically, 66% of the variance in VIQ at time 2 was accounted for by VIQ at time 1, a further 20% was accounted for by the change in grey matter density in the left motor speech region, with the remaining 14% unaccounted for. Similarly, 35% of the variance in PIQ at time 2 was accounted for by PIQ at time 1, with 13% accounted for by the change in grey matter density in the anterior cerebellum, leaving 52% unaccounted for. Future studies may be able to account for more of the between-subject variability by using a similar methodology with larger samples or other methodologies that measure structural or functional connectivity^{8,19}.

Our findings demonstrate considerable effects of brain plasticity in our sample during the teenage years, over and above normal development. By obviating the many sources of between-subject variance and controlling for global changes in brain structure, our within-subject analysis has allowed us to dissociate brain regions where structure reflects individual differences in verbal or non-verbal performance, in a way that has proved difficult in previous studies using behavioural data from a single point in time. We have also shown that the changes observed over time in the IQ scores of teenagers cannot simply be measurement error, because they correlate with independently measured changes in brain structure in regions that are plausibly related to the verbal and non-verbal functions tested. Further studies are required to determine the generalizability of this finding; for example, the same degree of plasticity may be present throughout life or the adolescent years covered by this study may be special in this regard. In addition, future work could consider the causes of the identified changes both in intelligence and in brain structure and how they impact on educational performance and employment prospects. The implication of our present findings is that an individual's strengths and weaknesses in skills relevant to education and employment are still emerging or changing in the teenage years.

METHODS SUMMARY

This study was approved by the Joint Ethics Committee of the Institute of Neurology and the National Hospital for Neurology and Neurosurgery, London, UK. All structural and functional scans at times 1 and 2 were acquired from the same Siemens 1.5T Sonata MRI scanner (Siemens Medical Systems). The structural images were acquired using a T1-weighted modified driven equilibrium Fourier transform sequence with 176 sagittal partitions and an image matrix of 256×224 , yielding a final resolution of 1 mm^3 (repetition time, 12.24 ms; echo time, 3.56 ms; inversion time, 530 ms). To pre-process the 66 structural images (33 participants \times 2 time points), we used SPM8 (<http://www.fil.ion.ucl.ac.uk/spm>) with the DARTEL toolbox to segment and spatially normalize the brains into the same template, with and without modulation. Modulated images incorporate a measure of local brain volume, whereas unmodulated images, used with proportional scaling to correct for global grey matter, provide a measure of regional grey matter density. Previous studies^{20–22} have shown that the correlations between brain structure and cognitive ability are better detected by grey matter density. Coordinates for each voxel were converted to standard MNI space. Images were smoothed using a Gaussian kernel with an isotropic full-width of 8 mm at half-maximum. The relationship between change in IQ and change in brain structure was investigated by entering the appropriate pre-processed images (modulated or unmodulated grey or white matter) into within-subject paired t -tests, with change in IQ (VIQ, PIQ or FSIQ) and year of scan as covariates. The degree to which IQ at time 2 was predicted by changes in brain structure was investigated in a hierarchical regression analysis with IQ at time 1 entered before change in brain structure. Details of the functional imaging method have been reported elsewhere^{23–25} and are summarized in Supplementary Information.

Received 17 May; accepted 26 August 2011.

Published online 19 October 2011.

- McCall, R. B. Childhood IQs as predictors of adult educational and occupational status. *Science* **197**, 482–483 (1977).
- Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R. & Starr, J. M. The stability of differences in mental ability from childhood to old age: follow-up of the 1932 Scottish Mental Survey. *Intelligence* **28**, 49–55 (2000).
- Wilke, M., Sohn, J.-H., Byars, A. W. & Holland, S. K. Bright spots: correlations of gray matter volume with IQ in a normal pediatric population. *Neuroimage* **20**, 202–215 (2003).
- Gong, Q.-Y. *et al.* Voxel-based morphometry and stereology provide convergent evidence of the importance of medial prefrontal cortex for fluid intelligence in healthy adults. *Neuroimage* **25**, 1175–1186 (2005).
- Camara, W. J., Nathan, J. S. & Puente, A. E. Psychological test usage: implications in professional psychology. *Prof. Psychol. Res. Pr.* **31**, 141–154 (2000).
- Kaufman, A. & Lichtenberger, E. O. *Assessing Adolescent and Adult Intelligence* 209–216 (Wiley, 2006).
- Haier, R. J., Jung, R. E., Yeo, R. A., Head, K. & Alkire, M. T. Structural brain variation and general intelligence. *Neuroimage* **23**, 425–433 (2004).
- Colom, R., Karama, S., Jung, R. E. & Haier, R. J. Human intelligence and brain networks. *Dialogues Clin. Neurosci.* **12**, 489–501 (2010).
- Shaw, P. *et al.* Intellectual ability and cortical development in children and adolescents. *Nature* **440**, 676–679 (2006).
- Huang, J., Carr, T. H. & Cao, Y. Comparing cortical activations for silent and overt speech using event-related fMRI. *Hum. Brain Mapp.* **15**, 39–53 (2002).

11. Nitschke, M. F., Kleinschmidt, A., Wessel, K. & Frahm, J. Somatotopic motor representation in the human anterior cerebellum: a high-resolution functional MRI study. *Brain* **119**, 1023–1029 (1996).
12. Stoodley, C. J., Valerad, E. M. & Schmahmann, J. D. An fMRI study of intra-individual functional topography in the human cerebellum. *Behav. Neurol.* **23**, 65–79 (2010).
13. Diamond, A. Close interrelation of motor development and cognitive development and of the cerebellum and prefrontal cortex. *Child Dev.* **71**, 44–56 (2000).
14. Pangelinan, M. M. *et al.* Beyond age and gender: relationships between cortical and subcortical brain volume and cognitive-motor abilities in school-age children. *Neuroimage* **54**, 3093–3100 (2011).
15. Davis, A. S., Pass, L. A., Finch, W. H., Dean, R. S. & Woodcock, R. W. The canonical relationship between sensory-motor functioning and cognitive processing in children with attention-deficit/hyperactivity disorder. *Arch. Clin. Neuropsychol.* **24**, 273–286 (2009).
16. Davis, E. E., Pitchford, N. J., Jaspan, T., McArthur, D. & Walker, D. Development of cognitive and motor function following cerebellar tumour injury sustained in early childhood. *Cortex* **46**, 919–932 (2010).
17. Rosenbaum, D. A., Carlson, R. A. & Gilmore, R. O. Acquisition of intellectual and perceptual-motor skills. *Annu. Rev. Psychol.* **52**, 453–470 (2001).
18. Wassenberg, R. *et al.* Relation between cognitive and motor performance in 5- to 6-year-old children: results from a large-scale cross-sectional study. *Child Dev.* **76**, 1092–1103 (2005).
19. Jung, R. E. & Haier, R. J. The parieto-frontal integration theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behav. Brain Sci.* **30**, 135–154 (2007).
20. Eckert, M. A. *et al.* To modulate or not to modulate: differing results in uniquely shaped Williams syndrome brains. *Neuroimage* **32**, 1001–1007 (2006).
21. Lee, H. *et al.* Anatomical traces of vocabulary acquisition in the adolescent brain. *J. Neurosci.* **27**, 1184–1189 (2007).
22. Richardson, F. M., Thomas, M. S., Filippi, R., Harth, H. & Price, C. J. Contrasting effects of vocabulary knowledge on temporal and parietal brain structure across lifespan. *J. Cogn. Neurosci.* **22**, 943–954 (2010).
23. Seghier, M. L., Fagan, E. & Price, C. J. Functional subdivisions in the left angular gyrus where the semantic system meets and diverges from the default network. *J. Neurosci.* **30**, 16809–16817 (2010).
24. Seghier, M. L. & Price, C. J. Dissociating functional brain networks by decoding the between-subject variability. *Neuroimage* **45**, 349–359 (2009).
25. Parker Jones, 'Ö. *et al.* Where, when and why brain activation differs for bilinguals and monolinguals during picture naming and reading aloud. *Cereb. Cortex* advance online publication, (<http://dx.doi.org/10.1093/cercor/bhr161>) (24 June 2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was funded by the Wellcome Trust. We thank J. Glensman, A. Brennan, A. Peters, L. Stewart, K. Pitcher and R. Rutherford for their help with data collection; and W. Penny for his advice on statistical analyses.

Author Contributions C.J.P. designed and supervised the study. C.J.P. and C.S. recruited the participants. C.S., S.R. and G.J. collected the data. F.M.R., S.R., C.E., M.L.S. and C.J.P. analysed the data. S.R., M.S.C.T. and C.J.P. wrote the manuscript and all authors edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.J.P. (c.price@fil.ion.ucl.ac.uk).

Non-canonical inflammasome activation targets caspase-11

Nobuhiko Kayagaki¹, Søren Warming², Mohamed Lamkanfi^{3,4}, Lieselotte Vande Walle^{3,4}, Salina Louie¹, Jennifer Dong¹, Kim Newton¹, Yan Qu¹, Jinfeng Liu⁵, Sherry Heldens², Juan Zhang⁶, Wyne P. Lee⁶, Merone Roose-Girma² & Vishva M. Dixit¹

Caspase-1 activation by inflammasome scaffolds comprised of intracellular nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs) and the adaptor ASC is believed to be essential for production of the pro-inflammatory cytokines interleukin (IL)-1 β and IL-18 during the innate immune response^{1–5}. Here we show, with C57BL/6 *Casp11* gene-targeted mice, that caspase-11 (also known as caspase-4)^{6–8} is critical for caspase-1 activation and IL-1 β production in macrophages infected with *Escherichia coli*, *Citrobacter rodentium* or *Vibrio cholerae*. Strain 129 mice, like *Casp11*^{–/–} mice, exhibited defects in IL-1 β production and harboured a mutation in the *Casp11* locus that attenuated caspase-11 expression. This finding is important because published targeting of the *Casp1* gene was done using strain 129 embryonic stem cells^{9,10}. *Casp1* and *Casp11* are too close in the genome to be segregated by recombination; consequently, the published *Casp1*^{–/–} mice lack both caspase-11 and caspase-1. Interestingly, *Casp11*^{–/–} macrophages secreted IL-1 β normally in response to ATP and monosodium urate, indicating that caspase-11 is engaged by a non-canonical inflammasome. *Casp1*^{–/–} *Casp11*^{129mt/129mt} macrophages expressing caspase-11 from a C57BL/6 bacterial artificial chromosome transgene failed to secrete IL-1 β regardless of stimulus, confirming an essential role for caspase-11 in IL-1 β production. Caspase-11 rather than caspase-1, however, was required for non-canonical inflammasome-triggered macrophage cell death, indicating that caspase-11 orchestrates both caspase-1-dependent and -independent outputs. Caspase-1 activation by non-canonical stimuli required NLRP3 and ASC, but caspase-11 processing and cell death did not, implying that there is a distinct activator of caspase-11. Lastly, loss of caspase-11 rather than caspase-1 protected mice from a lethal dose of lipopolysaccharide. These data highlight a unique pro-inflammatory role for caspase-11 in the innate immune response to clinically significant bacterial infections.

Many bacterial toxins promote inflammasome activation^{11–15}, and this is also true for cholera toxin B (CTB), a component of the AB5 holotoxin complex (Fig. 1a and Supplementary Fig. 1). CTB lacks the enzymatic activity necessary for entero-pathogenesis, but binds to GM1 ganglioside on the cell surface to facilitate the entry of enzymatic component cholera toxin A (CTA)¹⁶. Similar to ATP, which is a known activator of caspase-1 (ref. 17), CTB induced NLRP3- and ASC-dependent processing and secretion of caspase-1 and IL-1 β from lipopolysaccharide (LPS)-primed bone-marrow-derived C57BL/6 macrophages (BMDMs). Several other NLR family members (NLRC4, NLRP6, NLRP12, NOD1 and NOD2), inflammasome component AIM2 (refs 18–20), the ATP-releasing pannexin-1 channel, and the ATP-gated P2RX7 receptor²¹ were dispensable for CTB-induced IL-1 β secretion (Supplementary Fig. 1a). These data indicate that CTB does not activate the NLRP3 inflammasome indirectly by causing ATP release and P2RX7 receptor stimulation. To investigate

other host factors involved in the inflammasome response to CTB, we determined whether macrophages from other mouse strains secrete IL-1 β when exposed to CTB. Polymorphisms in the *Nlrp1b* gene, for example, alter sensitivity to anthrax lethal toxin¹². LPS-primed 129S6 BMDMs cultured with ATP processed caspase-1 and secreted IL-1 β , but they were unresponsive to CTB (Fig. 1b, c). Western blotting for pro-inflammatory adaptor proteins and caspases indicated that 129S6 macrophages were deficient in pro-caspase-11 isoforms⁷ p43 and p38 (Fig. 1d). Quantitative polymerase chain reaction with reverse transcription (RT-PCR) confirmed attenuated LPS-induced *Casp11* messenger RNA expression in 129S6 BMDMs (Fig. 1e). Caspase-11 is similar to caspase-1 (46%), orthologous to human caspase-4 and -5, and has been shown to interact with caspase-1 upon overexpression^{6–8}. Although caspase-11 was shown to be essential for LPS-induced IL-1 β secretion *in vivo*⁸, the role of caspase-11 in macrophage inflammasome signalling and its mechanism of activation remain unclear.

Primers that amplified the full-length *Casp11* mRNA expressed in C57BL/6 BMDMs after culture with either LPS or *E. coli* recovered only a truncated *Casp11* complementary DNA (Δ 110) from similarly treated 129S6 cells (Fig. 1f). Sequencing revealed that the 129S6 *Casp11* transcript lacked all sequence encoded by exon 7. Splicing of exon 6 to exon 8 creates a frame-shift after proline 304 and a stop codon occurred after 5 aberrant amino acids (Fig. 1g). Absence of exon 7 expression was confirmed in 129S6 BMDMs by quantitative RT-PCR for *Casp11* exon 7 (Supplementary Fig. 2a). Macrophages from three additional 129 substrains (129X1, 129S1 and 129P3) also lacked detectable caspase-11 protein, expressed *Casp11* Δ 110 mRNA, and produced negligible IL-1 β in response to CTB (Supplementary Fig. 2b–d). Sequencing of 129S1 genomic DNA identified a 5-bp deletion encompassing the exon 7 splice acceptor junction as the origin of the Δ 110 isoform (Fig. 1h). The caspase-11 antibody used in our experiments detected the truncated caspase-11 protein encoded by *Casp11* Δ 110 when it was overexpressed in 293T cells (Supplementary Fig. 2e) but failed to detect endogenous caspase-11 Δ 110 product in 129 BMDMs (Fig. 1d and Supplementary Fig. 2b), possibly due to nonsense-mediated mRNA decay (Fig. 1e and Supplementary Fig. 2a). Caspase-11 Δ 110 protein product lacks most of the small catalytic subunit so any protein that is expressed should be non-functional (Supplementary Fig. 2f).

We explored further the role of caspase-11 in inflammasome activation by targeting *Casp11* exon 5 for deletion in C57BL/6 embryonic stem (ES) cells. This exon encodes the critical catalytic residue cysteine 254 (ref. 7) and most of the caspase-11 large catalytic subunit (Supplementary Fig. 3). LPS-primed C57BL/6 *Casp11*^{+/+} (wild type) and *Casp11*^{–/–} BMDMs were stimulated with ATP to engage the NLRP3-dependent inflammasome, poly(dA:dT) double-stranded DNA or *Francisella tularensis* to activate the AIM2-dependent inflammasome, and flagellin or *Pseudomonas aeruginosa* to engage the NLRC4-dependent inflammasome^{18–20}. These canonical stimuli

¹Department of Physiological Chemistry, Genentech Inc., South San Francisco, California 94080, USA. ²Department of Molecular Biology, Genentech Inc., South San Francisco, California 94080, USA.

³Department of Medical Protein Research, VIB, B-9000 Ghent, Belgium. ⁴Department of Biochemistry, Ghent University, B-9000 Ghent, Belgium. ⁵Department of Bioinformatics, Genentech Inc., South San Francisco, California 94080, USA. ⁶Department of Immunology, Genentech Inc., South San Francisco, California 94080, USA.

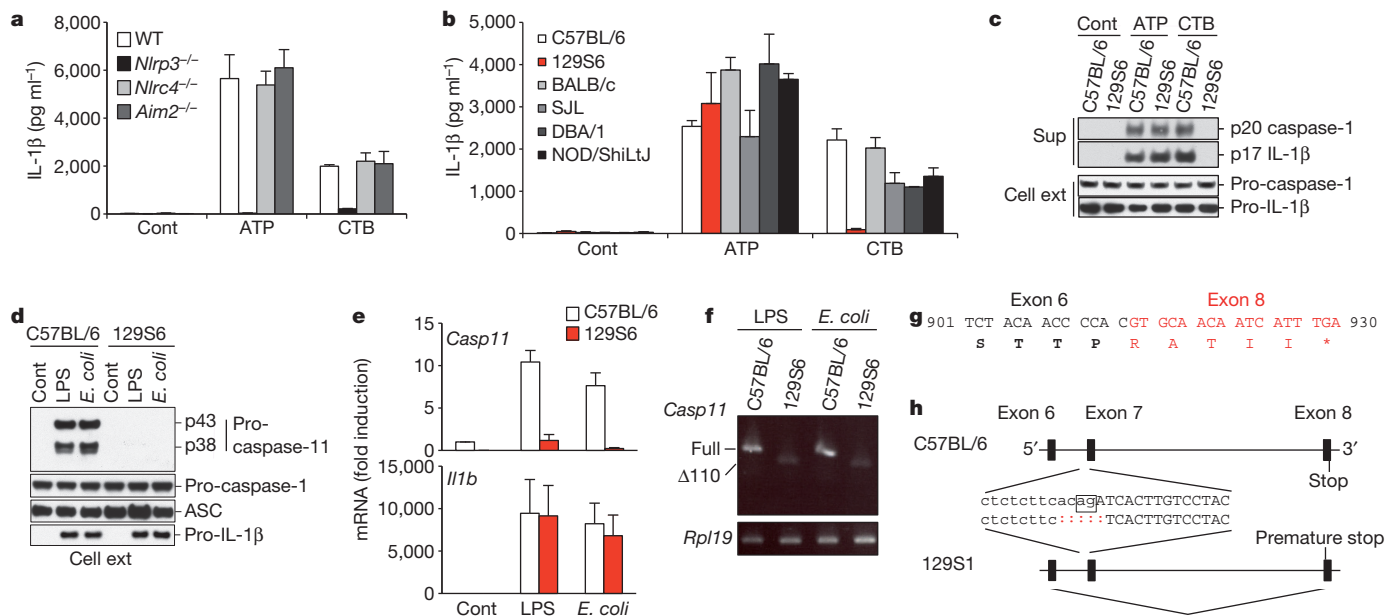


Figure 1 | Germline mutation of *Casp11* in mouse strain 129 abolishes inflammasome activation by CTB. **a**, **b**, IL-1 β secretion by LPS-primed BMDMs after culture in medium alone (Cont), 5 mM ATP or 20 μ g ml $^{-1}$ CTB for 16 h. WT, wild type. **c**, LPS-primed BMDMs stimulated for 8 h. IL-1 β and the caspase-1 p20 subunit were immunoblotted in cell supernatants (Sup), pro-caspase-1 and pro-IL-1 β in cell extracts (Cell ext). **d**, BMDM extracts after 6 h.

induced comparable IL-1 β secretion from wild-type and *Casp11* $^{-/-}$ BMDMs (Fig. 2a). Like *Casp11* mutant 129 BMDMs, however, *Casp11* $^{-/-}$ BMDMs failed to secrete IL-1 β in response to CTB (Fig. 2a). IL-1 β secretion in response to other bacterial toxins, including adenylyl cyclase (AC) toxin¹⁴, listeriolysin O (LLO) toxin¹⁵ or *Clostridium difficile* toxin B¹³, was not affected by caspase-11 deficiency. Caspase-11 also was dispensable for NLRP3-dependent IL-1 β secretion in response to monosodium urate (MSU), calcium pyrophosphate (CPPD)^{2,22} or the ionophore nigericin¹⁷ (Fig. 2b). In contrast, NLRP3- and ASC-dependent IL-1 β secretion from BMDMs infected with live *E. coli*, *C. rodentium* and *V. cholerae*²³, required caspase-11, with or without LPS priming (Fig. 2b and Supplementary Fig. 4a–c). Hereafter, we refer to this caspase-11-dependent inflammasome as the non-canonical inflammasome, and all stimulations are performed with LPS-primed BMDMs. The non-canonical inflammasome that we describe seems to be distinct from the human caspase-5/NLRP1 complex reported previously²⁴ because *Casp11*^{129mt/129mt} BMDMs exposed to anthrax lethal toxin secreted IL-1 β normally²⁵ (data not shown).

Caspase-1 is thought to cleave pro-IL-1 β and pro-IL-18 into biologically active IL-1 β p17 and IL-18 p18, respectively^{5,9,10,26,27}. Secretion of processed caspase-1 subunits p20 and p10, IL-1 β and IL-18 in response to CTB or *E. coli* required caspase-11, ASC and NLRP3 (Fig. 2c, d). Wild-type BMDMs cleaved pro-caspase-11 and produced the autocatalytically cleaved caspase-11 p26 subunit^{6,7} (Supplementary Fig. 4d) in response to CTB or *E. coli* (Fig. 2e). ATP also stimulated formation of the caspase-11 p26 subunit (Fig. 2e) but some time after caspase-1 processing peaked (data not shown). Because caspase-11 was dispensable for ATP-induced processing and secretion of caspase-1, IL-1 β and IL-18 (Fig. 2a, c, d and Supplementary Fig. 4f), we believe that this caspase-11 processing after ATP treatment reflects the promiscuity of caspase-1 (ref. 28) in our *in vitro* assay conditions and does not represent a physiological function of caspase-11. Consistent with a role for caspase-11 in non-canonical caspase-1 processing, endogenous pro-caspase-11 p43 and its p26 cleavage product co-immunoprecipitated with endogenous caspase-1 from BMDMs treated with CTB or *E. coli* (Supplementary Fig. 4e). A much weaker interaction occurred between pro-caspase-1 and pro-caspase-11 after

e, *Casp11* and *Il1b* mRNA expression in BMDMs after 6 h. **f**, Full-length *Casp11* transcripts recovered from BMDMs in **e**. **g**, 129 *Casp11* Δ 110 isoform splices exon 6 (black) to exon 8 (red). **h**, 129S1 and C57BL/6 *Casp11* genes. Exon 7 sequence is uppercase and the splice acceptor sequence is boxed. Graphs show the mean \pm s.d. of triplicate wells and are representative of three independent experiments.

canonical ATP treatment. Unexpectedly, CTB and *E. coli* both induced secretion of the caspase-11 p26 subunit in the absence of NLRP3 or ASC (Fig. 2e). The mechanism of caspase-11 activation remains unknown.

Canonical inflammasome activation by ATP causes NLRP3- and ASC-dependent macrophage death and releases pro-inflammatory IL-1 α and high-mobility group box 1 protein (HMGB1), which both lack a leader sequence and a typical caspase cleavage site^{29,30} (Fig. 2d, e). Caspase-11 was not required for this death because wild-type and *Casp11* $^{-/-}$ BMDMs stimulated with either ATP or *C. difficile* toxin B released comparable amounts of HMGB1 (Fig. 2e), IL-1 α and an indicator of cell membrane damage, lactate dehydrogenase (LDH; Fig. 2d). In contrast, caspase-11 deficiency, rather than NLRP3 or ASC loss, prevented HMGB1, IL-1 α and LDH release triggered by non-canonical activators CTB, *E. coli*, *C. rodentium* and *V. cholerae* (Fig. 2d, e). Collectively, our data indicate that caspase-11 is an essential effector of non-canonical caspase-1 activation and IL-1 β /IL-18 secretion together with NLRP3 and ASC, but signals death independently of NLRP3 and ASC (Fig. 2f).

Next we assessed non-canonical inflammasome activation in *Casp1* $^{-/-}$ mice¹⁰, but this strain and an independent knockout strain⁹ were generated from 129 ES cells harbouring the *Casp11* mutation described in Fig. 1. *Casp1* and *Casp11* are adjacent in the mouse genome and separated by \sim 1,500 base pairs. Consequently, backcrossing to C57BL/6 or NOD/ShiLtJ mice over multiple generations is highly unlikely to segregate the two mutant genes. Indeed, *Casp1* $^{-/-}$ BMDMs expressed the *Casp11* Δ 110 isoform and lacked detectable caspase-11 protein after stimulation with diverse Toll-like receptor agonists regardless of their genetic background⁶ (Fig. 3a, b and Supplementary Fig. 5a, b). Hereafter, we therefore refer to these *Casp1* $^{-/-}$ *Casp11*^{129mt/129mt} mice as *Casp1/11* double-knockout mice.

To obtain mice lacking only caspase-1, we microinjected a *Casp11* bacterial artificial chromosome (BAC) transgene (Tg) into *Casp1/11* double-knockout embryos. *Casp1* exons 1 and 2 within the BAC were deleted to prevent any caspase-1 expression. Western blotting identified a *Casp1* $^{-/-}$ *Casp11*^{Tg} line that exhibited inducible caspase-11 expression (Fig. 3c), and expression of *Casp11* exon 7 was confirmed

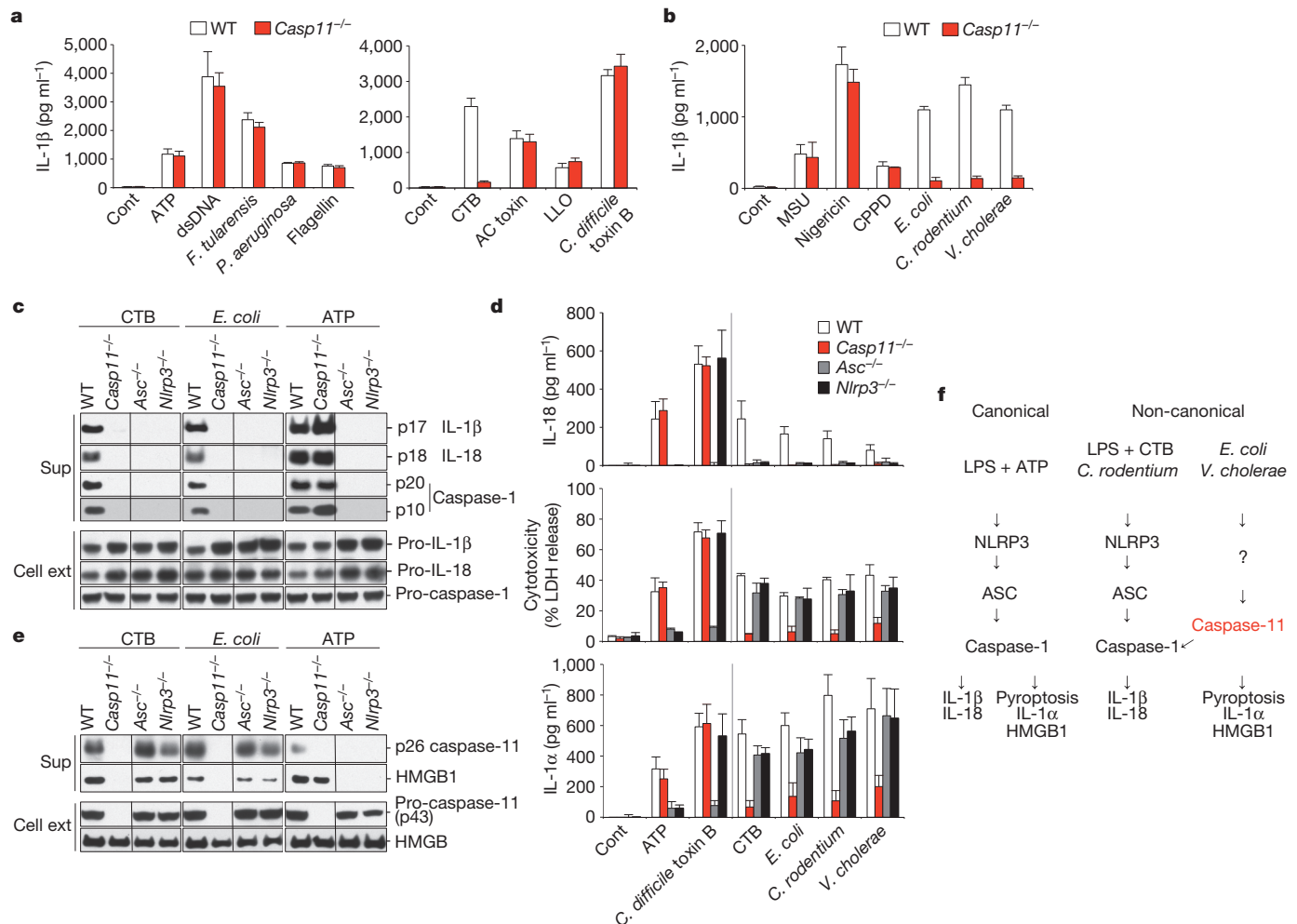


Figure 2 | Caspase-11 mediates non-canonical inflammasome activation by CTB, *E. coli*, *C. rodentium* and *V. cholerae*. **a, b,** IL-1β secretion by LPS-primed BMDMs stimulated as indicated for 16 h. Cont, medium alone. **c, e,** Immunoblots of IL-1β, IL-18, cleaved caspase-1 subunits (**c**), caspase-11 p26 and HMGB1 (**e**) released from LPS-primed BMDMs after stimulation for

8 h. **d,** IL-18, LDH and IL-1α released from LPS-primed BMDMs after stimulation for 16 h (ATP, for 8 h). **f,** Model for canonical and non-canonical inflammasome signalling. Graphs show the mean \pm s.d. of triplicate wells and are representative of three independent experiments.

by quantitative RT-PCR (Supplementary Fig. 6a). *Casp1*^{-/-}*Casp11*^{Tg} or *Casp1/11* double-knockout BMDMs, unlike their wild-type counterparts, both failed to process and secrete IL-1β and IL-18 in response to ATP, *C. difficile* toxin B, CTB or *E. coli* (Fig. 3d, e and Supplementary Fig. 6b, c). These data confirm the essential role of caspase-1 in IL-1β and IL-18 production. Both caspase-1 and caspase-11 were required for IL-1β and IL-18 secretion triggered by non-canonical stimuli, and although recombinant caspase-11 cleaved caspase-1, it processed IL-1β poorly in the absence of caspase-1 (refs 6, 7; data not shown). Therefore, pro-IL-1β probably is a direct substrate of caspase-1 rather than caspase-11 in the non-canonical inflammasome pathway.

Next we determined the contribution of caspase-1 to macrophage death induced by non-canonical activators CTB and *E. coli* (Fig. 3f). *Casp11*^{+/-} and *Casp1*^{-/-}*Casp11*^{Tg} showed similar LDH release to wild-type BMDMs, whereas *Casp1/11* double-knockout and *Casp11*^{-/-} BMDMs showed enhanced survival. The *Casp11* transgene also restored IL-1α (Fig. 3f) and HMGB1 release (Fig. 3g) from *Casp1/11* double-knockout BMDMs, with caspase-11 p26 being released as in wild-type BMDMs (Fig. 3g). These results indicate that caspase-1 is dispensable for caspase-11 activation and macrophage death induced by the non-canonical inflammasome (Fig. 2f). In contrast, macrophage death plus the release of IL-1α and HMGB1 induced by canonical activators ATP and *C. difficile* toxin B required caspase-1 but not caspase-11 (Fig. 3f, g). The molecular mechanisms distinguishing

caspase-1 activation in the canonical versus non-canonical pathways remains to be determined.

We extended our findings with cultured macrophages by challenging wild type, *Casp11*^{+/-}, *Casp11*^{-/-}, *Casp1/11* double knockout and *Casp1*^{-/-}*Casp11*^{Tg} with a lethal dose of LPS (54 mg kg⁻¹), which is a model of acute septic shock (Fig. 4a and Supplementary Table 1). Consistent with previous reports^{8,10}, *Casp11*^{-/-} and *Casp1/11* double-knockout mice were resistant to LPS challenge; whereas six of six *Casp11*^{-/-} mice were alive at 40 h after injection, six of six wild-type and *Casp11*^{+/-} mice had succumbed within 18 h. Notably, no *Casp1*^{-/-}*Casp11*^{Tg} mice survived beyond 26 h. We conclude that caspase-11 rather than caspase-1 is the dominant effector of LPS-induced lethal septic shock. Previous studies implicating both caspase-11 and caspase-1 in this disease model^{8,10} may have been misled by the assumption that published *Casp1*^{-/-} mice expressed caspase-11 normally.

Casp11^{-/-}, *Casp1/11* double-knockout and *Casp1*^{-/-}*Casp11*^{Tg} mice all had less serum IL-1β and IL-18 at 12 h after LPS injection than wild-type or *Casp11*^{+/-} mice, indicating that both caspase-1 and caspase-11 were necessary for IL-1β and IL-18 secretion *in vivo*^{8,10,26,27} (Fig. 4b). It should be noted that IL-1β and IL-18 are dispensable for LPS-induced lethality because mice lacking both IL-1β and IL-18 are as susceptible as wild-type mice³⁰. We speculate that caspase-11-mediated tissue damage, which probably does not require caspase-1, is responsible

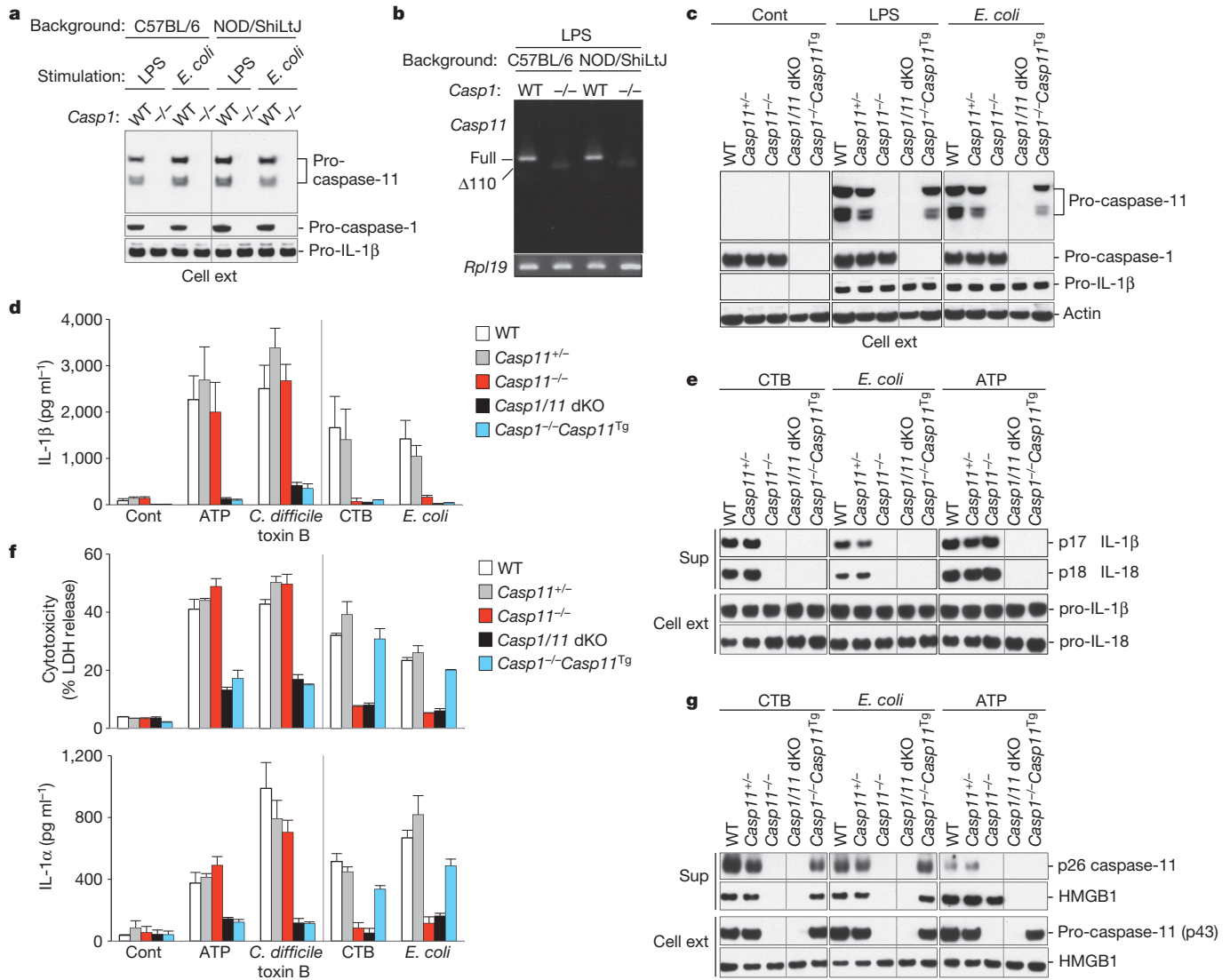


Figure 3 | Caspase-1 and caspase-11 have stimulus-specific roles during inflammasome activation. **a–c**, BMDMs stimulated as indicated for 6 h were immunoblotted (**a**, **c**) or *Casp11* transcripts were recovered by RT-PCR (**b**). Cont, medium alone. dKO, double knockout. **d**, IL-1 β secretion by LPS-primed BMDMs stimulated for 16 h. **e**, LPS-primed BMDMs stimulated for 8 h. IL-1 β and IL-18 were immunoblotted in cell supernatants (Sup), pro-IL-1 β and

pro-IL-18 in cell extracts (Cell ext). **f**, LDH and IL-1 α released from LPS-primed BMDMs after stimulation for 16 h (ATP, for 8 h). **g**, LPS-primed BMDMs stimulated for 8 h. Caspase-11 p26 and HMGB1 were immunoblotted in cell supernatants, pro-caspase-11 and HMGB1 in cell extracts. Graphs show the mean \pm s.d. of triplicate wells and are representative of three independent experiments.

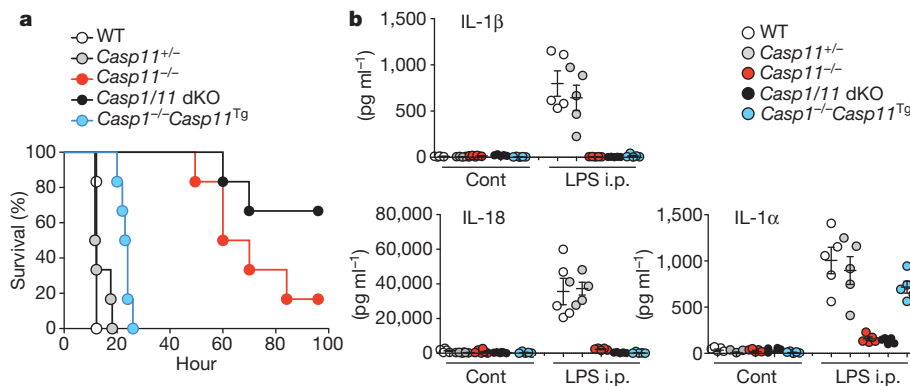


Figure 4 | Caspase-11 rather than caspase-1 is required for LPS-induced lethality. **a**, Survival of mice ($n = 6$ per genotype) injected intraperitoneally (i.p.) with 54 mg kg $^{-1}$ LPS. Data are representative of two independent experiments. Adjusted *P* values are supplied in Supplementary Table 1.

b, Serum IL-1 β , IL-18 and IL-1 α at 12 h after injection of 20 mg kg $^{-1}$ LPS. Circles represent individual mice. Error bars represent the mean \pm s.e.m. Data are representative of two independent experiments.

for the lethal septic shock. Intriguingly, IL-1 α released from damaged cells into the serum was equivalent in wild-type, *Casp11*^{+/−} and *Casp1*^{−/−}*Casp11*^{Tg} mice, whereas very little IL-1 α was detected in *Casp11*^{−/−} or *Casp1/11* double-knockout mice (Fig. 4b). Mice lacking the IL-1 receptor are as susceptible as wild-type mice to LPS-induced lethal sepsis (data not shown), so it is not clear what causes the lethality downstream of caspase-11. Like *Casp1*^{−/−}*Casp11*^{Tg} mice, *Nlrp3*^{−/−} and *Asc*^{−/−} mice survived only slightly longer than wild-type mice following high dose LPS (Supplementary Fig. 7a), despite failing to make serum IL-1 β and IL-18 (Supplementary Fig. 7b). This result confirms previous studies showing that NLRP3 and ASC are essential for LPS-induced caspase-1 activation *in vivo*^{17,29}, and it supports the notion that the caspase-1-independent non-canonical pathway elicits lethal septic shock. The NLRP3–ASC–caspase-1 axis may amplify the shock response, however, as *Nlrp3*^{−/−} or *Asc*^{−/−} mice were relatively resistant to lower doses of LPS^{17,18}.

Pro-inflammatory caspase-11 triggers caspase-1-independent macrophage death and caspase-1-dependent IL-1 β and IL-18 production in response to a subset of inflammasome activators that we refer to as non-canonical activators. Our *in vivo* data indicate, contrary to current thinking, that caspase-11 rather than caspase-1 may be the critical effector of deleterious inflammatory responses. Therefore, targeting human caspase-4 and caspase-5 may be more effective than caspase-1 inhibition in patients with sepsis. Our findings also highlight the need to revisit the role of caspase-1 versus caspase-11 in different mouse disease models, as so far all studies have used *Casp1/11* double-knockout mice.

METHODS SUMMARY

Mice. *Casp11* exon 5 in C57BL/6 ES cells was flanked by *loxP* sites and then deleted with Cre recombinase. *Casp1*^{−/−}*Casp11*^{Tg} mice were created by introducing a *Casp11* BAC transgene into *Casp1/11* double-knockout embryos.

BMDM culture. BMDMs were primed with 500 ng ml^{−1} LPS (*E. coli* 0111:B4) for 5 h and then stimulated in OPTI-MEM with 500 ng ml^{−1} LPS or infected with bacteria. IL-1 β , IL-1 α and IL-18 secretion was measured by ELISA. For immunoblotting, cells were lysed with RIPA buffer. Proteins in culture supernatants were precipitated with 7.2% trichloroacetic acid plus 0.15% sodium cholate.

LPS-primed BMDMs were infected with *P. aeruginosa* (multiplicity of infection (m.o.i.) 25), *E. coli* (m.o.i. 20), *C. rodentium* (m.o.i. 20), *V. cholerae* (m.o.i. 50) or *F. tularensis* ssp. *novi* strain U112 (m.o.i. 40) for 1.5 h and then cultured in 100 μ g ml^{−1} gentamycin. Other stimulations included ATP (5 mM), MSU (100 μ g ml^{−1}), CPPD (200 μ g ml^{−1}), CTB (20 μ g ml^{−1}), AC toxin (5 μ g ml^{−1}), *C. difficile* toxin B (0.2 μ g ml^{−1}) and LLO toxin (10 μ g ml^{−1}).

Endotoxic shock model. Mice (8–10 weeks old) were injected intraperitoneally with 54 mg kg^{−1} LPS (*E. coli* 0111:B4) and monitored 8 times daily for a total of 6 days. For serum cytokine measurements, a separate cohort of mice received 20 mg kg^{−1} LPS and blood was collected 12 h later.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 July; accepted 13 September 2011.

Published online 16 October 2011.

1. Thornberry, N. A. *et al.* A novel heterodimeric cysteine protease is required for interleukin-1 beta processing in monocytes. *Nature* **356**, 768–774 (1992).
2. Schroder, K. & Tschopp, J. The inflammasomes. *Cell* **140**, 821–832 (2010).
3. Jin, C. & Flavell, R. A. Molecular mechanism of NLRP3 inflammasome activation. *J. Clin. Immunol.* **30**, 628–631 (2010).
4. Franchi, L., Warner, N., Viani, K. & Nunez, G. Function of Nod-like receptors in microbial recognition and host defense. *Immunol. Rev.* **227**, 106–128 (2009).
5. Dinarello, C. A. Interleukin-1 in the pathogenesis and treatment of inflammatory diseases. *Blood* **117**, 3720–3732 (2011).
6. Kang, S. J. *et al.* Dual role of caspase-11 in mediating activation of caspase-1 and caspase-3 under pathological conditions. *J. Cell Biol.* **149**, 613–622 (2000).

7. Wang, S. *et al.* Identification and characterization of Ich-3, a member of the interleukin-1 β converting enzyme (ICE)/Ced-3 family and an upstream regulator of ICE. *J. Biol. Chem.* **271**, 20580–20587 (1996).
8. Wang, S. *et al.* Murine caspase-11, an ICE-interacting protease, is essential for the activation of ICE. *Cell* **92**, 501–509 (1998).
9. Kuida, K. *et al.* Altered cytokine export and apoptosis in mice deficient in interleukin-1 β converting enzyme. *Science* **267**, 2000–2003 (1995).
10. Li, P. *et al.* Mice deficient in IL-1 β -converting enzyme are defective in production of mature IL-1 β and resistant to endotoxic shock. *Cell* **80**, 401–411 (1995).
11. Freche, B., Reig, N. & van der Goot, F. G. The role of the inflammasome in cellular responses to toxins and bacterial effectors. *Semin. Immunopathol.* **29**, 249–260 (2007).
12. Boyden, E. D. & Dietrich, W. F. *Nalp1b* controls mouse macrophage susceptibility to anthrax lethal toxin. *Nature Genet.* **38**, 240–244 (2006).
13. Ng, J. *et al.* *Clostridium difficile* toxin-induced inflammation and intestinal injury are mediated by the inflammasome. *Gastroenterology* **139**, 542–552 (2010).
14. Dunne, A. *et al.* Inflammasome activation by adenylate cyclase toxin directs Th17 responses and protection against *Bordetella pertussis*. *J. Immunol.* **185**, 1711–1719 (2010).
15. Meixenberger, K. *et al.* *Listeria monocytogenes*-infected human peripheral blood mononuclear cells produce IL-1 β , depending on listeriolysin O and NLRP3. *J. Immunol.* **184**, 922–930 (2010).
16. Beddoe, T., Paton, A. W., Le Nours, J., Rossjohn, J. & Paton, J. C. Structure, biological functions and applications of the AB5 toxins. *Trends Biochem. Sci.* **35**, 411–418 (2010).
17. Mariathasan, S. *et al.* Cryopyrin activates the inflammasome in response to toxins and ATP. *Nature* **440**, 228–232 (2006).
18. Mariathasan, S. *et al.* Differential activation of the inflammasome by caspase-1 adaptors ASC and Ipaf. *Nature* **430**, 213–218 (2004).
19. Fernandes-Alnemri, T., Yu, J. W., Datta, P., Wu, J. & Alnemri, E. S. AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA. *Nature* **458**, 509–513 (2009).
20. Hornung, V. *et al.* AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. *Nature* **458**, 514–518 (2009).
21. Solle, M. *et al.* Altered cytokine production in mice lacking P2X₇ receptors. *J. Biol. Chem.* **276**, 125–132 (2001).
22. Martinon, F., Petrilli, V., Mayor, A., Tardivel, A. & Tschopp, J. Gout-associated uric acid crystals activate the NALP3 inflammasome. *Nature* **440**, 237–241 (2006).
23. Toma, C. *et al.* Pathogenic *Vibrio* activate NLRP3 inflammasome via cytotoxins and TLR/nucleotide-binding oligomerization domain-mediated NF- κ B signaling. *J. Immunol.* **184**, 5287–5297 (2010).
24. Martinon, F., Burns, K. & Tschopp, J. The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-1 β . *Mol. Cell* **10**, 417–426 (2002).
25. Wickliffe, K. E., Leppla, S. H. & Moayeri, M. Anthrax lethal toxin-induced inflammasome formation and caspase-1 activation are late events dependent on ion fluxes and the proteasome. *Cell. Microbiol.* **10**, 332–343 (2008).
26. Ghayur, T. *et al.* Caspase-1 processes IFN- γ -inducing factor and regulates LPS-induced IFN- γ production. *Nature* **386**, 619–623 (1997).
27. Gu, Y. *et al.* Activation of interferon- γ inducing factor mediated by interleukin-1 β converting enzyme. *Science* **275**, 206–209 (1997).
28. Walsh, J. G., Logue, S. E., Luthi, A. U. & Martin, S. J. Caspase-1 promiscuity is counterbalanced by rapid inactivation of processed enzyme. *J. Biol. Chem.* **286**, 32513–32524 (2011).
29. Sutterwala, F. S. *et al.* Critical role for NALP3/CIA1/Cryopyrin in innate and adaptive immunity through its regulation of caspase-1. *Immunity* **24**, 317–327 (2006).
30. Lamkanfi, M. *et al.* Inflammasome-dependent release of the alarmin HMGB1 in endotoxemia. *J. Immunol.* **185**, 4385–4392 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank F.-X. Blaudin de Thé, A. Paler Martinez, R. J. Newman, X. Raidan, N. Ota, J. Ngo, L. Nguyen, A. Leung, L. Tam, M. Schlatter, H. Nguyen, V. Asghari and K. O'Rourke for technical support, M. van Lookeren Campagne, D. French, S. Mariathasan, T.-D. Kanneganti and D.M. Monack for discussion and reagents.

Author Contributions N.K., M.L., L.V.W., S.L., J.D., Y.Q. and S.H. designed and performed *in vitro* experiments. N.K., S.L., J.D., J.Z. and W.P.L. designed and performed *in vivo* experiments. S.W., M.R.-G. and K.N. generated the *Casp11*^{−/−} and *Casp1*^{−/−}*Casp11*^{Tg} mice. J.L. performed bioinformatics analyses. N.K., S.W., K.N. and V.M.D. prepared the manuscript. N.K. and V.M.D. contributed to the study design and data analyses.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.K. (kayagaki@gene.com) or V.M.D. (dixit@gene.com).

METHODS

Mice. *Nlrp3*^{-/-}, *Nlrp4*^{-/-}, *Nlrp12*^{-/-}, *Asc*^{-/-}, *Aim2*^{-/-}, *P2rx7*^{-/-} and *Panx1*^{-/-} mice were described previously^{31,32} and were backcrossed to C57BL/6 for at least 10 generations. Seven-to-ten-week-old male or female mice were used for experiments. Other mice included 129X1/SvJ (129X1), 129S1/SvImJ (129S1), 129P3/J (129P3), BALB/c, SJL, DBA/1 and NOD/ShiLtJ *Casp1/11* double knockout (Jackson Laboratory), 129S6/SvEvTac (129S6, Taconic), *Nlrp6*^{-/-}, *Nod1*^{-/-} and *Nod1*^{-/-}*Nod2*^{-/-} double knockout (Thirumala-Devi Kanneganti, St. Jude Children's Research Hospital). C57BL/6 *Casp1/11* double-knockout (*Casp1*^{-/-}*Casp11*^{129mt/129mt}) mice were obtained by backcrossing NOD/ShiLtJ *Casp1/11* double knockout to C57BL/6 for more than 10 generations. The Genentech animal care and use committee approved all mouse studies. *Casp1*^{-/-} mice were genotyped with PCR primers (5'TGAAATGCATGTACTGAGAGCAAGG; 5'CAATTGACCTGGGATTCTGG and 5'GTCAGAGATGAAAGACTTTGCTGC) yielding a 475-bp wild-type DNA fragment (a 704-bp fragment is possible too) and a 337-bp mutant DNA fragment.

Reagents. MSU, CPPD, flagellin (*S. typhimurium*) and ultra-pure LPS (*E. coli* 0111:B4) were from Invivogen. Other reagents included CTB, AC toxin, *C. difficile* toxin B, LLO toxin (List Biological Laboratories), Nigericin, poly(dA:dT) and ATP (Sigma).

BMDM culture. Bone marrow cells were differentiated in DMEM with 10% endotoxin-free fetal bovine serum (Omega Scientific) and 20% M-CSF-conditioned medium for 5–6 days, then plated at $\sim 1.0 \times 10^6$ cells ml⁻¹ and cultured overnight. BMDMs were primed with 500 ng ml⁻¹ LPS for 5 h and then cultured in OPTI-MEM media (Invitrogen) containing 500 ng ml⁻¹ LPS or bacteria (*P. aeruginosa*, ATCC 10145; *E. coli*, ATCC 11775; *C. rodentium*, ATCC 51116; *V. cholera*, ATCC 9459; *F. tularensis* ssp. *novidica* strain U112 from D. Monack, Stanford University). IL-1 β (Meso), IL-1 α (BD Bioscience) and IL-18 (MBL) were measured by ELISA. Cytotoxicity was measured by CytoTox 96 Non-Radioactive Cytotoxicity Assay (Promega).

Immunoblotting. Caspase-1 was immunoblotted with rat anti-mouse caspase-1 p20 (clone 4B4, Genentech) or rabbit anti-caspase-1 p10 (sc-514, Santa Cruz). A rabbit polyclonal antibody (GeneTex) detected IL-1 β . Other antibodies included rat anti-caspase-11 (clone 17D9, Novus Biologicals), rabbit anti-IL-18 (Biovision), rat anti-ASC (clone 8E4.1, Genentech), rabbit anti-HMGB1 (GeneTex) and rabbit anti- β -actin (Novus Biologicals).

RT-PCR. BMDM total RNA was prepared with an RNeasy kit (QIAGEN). The entire *Casp11* coding region was amplified with a SuperScript One-Step RT-PCR for Long Templates kit (Invitrogen). Primers were: 5'ATGGCTGAAAACAAACACCT and 5'TCAGTTGCCAGGAAAGAGGTAG. Primer and probe sets used for TaqMan (Applied Biosystems) included *Il1b* 171 (sense 5'GAGTGTGGATCCCAAGCAAT, anti-sense 5'TACAGTTGGGGAACTCTGC, probe 5'FAM-TGGAAAAACGGTTTGTCTTCA-TAMRA), *Casp11* 174 (sense 5'ACAATGCTGAACGCAGTGAC, anti-sense 5'CTGGTCTCTCCATTCCCA

GA, probe 5'FAM-CATTCTTCAGTGTGGACCCA-TAMRA), and *Rpl19* (sense 5'GCGCATCCTCATGGAGCACA, anti-sense GGTCAGCCAGGAGCTTCTTG, probe 5'FAM-CACAAGCTGAAGGCAGACAAGGCC C-TAMRA). Samples were normalized by quantification of *Rpl19* mRNA.

DNA sequencing. *Casp11* cDNAs amplified from C57BL/6 and 129S6 BMDMs were cloned into pcDNA3.1+ (Invitrogen) and sequenced. The caspase-11 C254A mutant was created with a QuikChange site-directed mutagenesis kit (Stratagene). *Casp11* genome sequence in 129S1 BAC clone (CT7-292K12; Invitrogen) was sequenced with primer 5'CATGTCTAACTATATTGAAATGTGAA.

BAC transgenic mouse. A C57BL/6 mouse BAC clone (RP23-78A8, Invitrogen) containing 156 kb of the *Casp1* and *Casp11* genomic region was characterized by DNA fingerprinting and transformed into SW102 cells. This BAC contains 65 kb DNA upstream of *Casp11* and 53 kb downstream of *Casp11*. To create a BAC transgene with functional *Casp11* and inactivated *Casp1*, an Frt-PGK-em7-Neo-Frt cassette was synthesized (Blue Heron/Origene) and used to delete, using recombinering, a 1.4-kb region encompassing *Casp1* exons 1 and 2 (NCBI37/mm9 assembly chr9: 5,298,191–5,299,550). Correctly targeted BAC DNA was transformed into Flp-expressing SW105 cells^{33,34} to remove the Neo selection marker. The modified region of the BAC was confirmed by DNA sequencing. Transgenic mice carrying the modified *Casp1/Casp11* BAC were obtained using standard pronuclear microinjection methods³⁵ using *Casp1/11* double-knockout embryos. *Casp1*^{-/-}*Casp11*^{Tg} mice were genotyped with PCR primers (5'ACAGAGAGATCTGAGCCTTCA and 5'ACACAGACTTGGACCCTGTAGTAG) yielding a 379 bp *Casp11*^{Tg} DNA fragment.

Endotoxin shock model. Mice (8–10 weeks old) were injected intraperitoneally with 54 mg kg⁻¹ LPS (*E. coli* 0111:B4; Sigma) and monitored 8 times daily for a total of 6 days. For serum cytokine measurements, a separate cohort of mice received 20 mg kg⁻¹ LPS and blood was collected 12 h later. Serum cytokines were measured by IL-1 α ELISA (BD Bioscience), IL-1 β ELISA (Meso) and IL-18 luminex assay (Bio-Rad).

- Jones, J. W. *et al.* Absent in melanoma 2 is required for innate immune recognition of *Francisella tularensis*. *Proc. Natl Acad. Sci. USA* **107**, 9771–9776 (2010).
- Qu, Y. *et al.* Pannexin-1 is required for ATP release during apoptosis but not for inflammasome activation. *J. Immunol.* **186**, 6553–6561 (2011).
- Warming, S., Costantino, N., Court, D. L., Jenkins, N. A. & Copeland, N. G. Simple and highly efficient BAC recombinering using galK selection. *Nucleic Acids Res.* **33**, e36 (2005).
- Lee, E. C. *et al.* A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**, 56–65 (2001).
- Van Keuren, M. L., Gavrilina, G. B., Filipiak, W. E., Zeidler, M. G. & Saunders, T. L. Generating transgenic mice from bacterial artificial chromosomes: transgenesis efficiency, integration and expression outcomes. *Transgenic Res.* **18**, 769–785 (2009).

Macrophage skewing by *Phd2* haploodeficiency prevents ischaemia by inducing arteriogenesis

Yukiji Takeda^{1,2,3,4}, Sandra Costa^{1,2,5*}, Estelle Delamarre^{1,2*}, Carmen Roncal^{1,2,3,4,6*}, Rodrigo Leite de Oliveira^{1,2,3,4}, Mario Leonardo Squadrito^{7,8}, Veronica Finisguerra^{1,2}, Sofie Deschoemaeker^{1,2}, Françoise Bruyère^{3,4}, Mathias Wenes^{1,2}, Alexander Hamm^{1,2}, Jens Serneels^{1,2}, Julie Magat⁹, Tapan Bhattacharyya^{10†}, Andrey Anisimov¹¹, Benedicte F. Jordan⁹, Kari Alitalo¹¹, Patrick Maxwell¹⁰, Bernard Gallez⁹, Zhen W. Zhuang¹², Yoshihiko Saito¹³, Michael Simons¹², Michele De Palma^{7†} & Massimiliano Mazzone^{1,2}

PHD2 serves as an oxygen sensor that rescues blood supply by regulating vessel formation and shape in case of oxygen shortage^{1–5}. However, it is unknown whether PHD2 can influence arteriogenesis. Here we studied the role of PHD2 in collateral artery growth by using hindlimb ischaemia as a model, a process that compensates for the lack of blood flow in case of major arterial occlusion^{6–8}. We show that *Phd2* (also known as *Egln1*) haplodeficient (*Phd2*^{+/-}) mice displayed preformed collateral arteries that preserved limb perfusion and prevented tissue necrosis in ischaemia. Improved arteriogenesis in *Phd2*^{+/-} mice was due to an expansion of tissue-resident, M2-like macrophages^{9,10} and their increased release of arteriogenic factors, leading to enhanced smooth muscle cell (SMC) recruitment and growth. Both chronic and acute deletion of one *Phd2* allele in macrophages was sufficient to skew their polarization towards a pro-arteriogenic phenotype. Mechanistically, collateral vessel preconditioning relied on the activation of canonical NF-κB pathway in *Phd2*^{+/-} macrophages. These results unravel how PHD2 regulates arteriogenesis and artery homeostasis by controlling a specific differentiation state in macrophages and suggest new treatment options for ischaemic disorders.

To understand whether partial loss of PHD2 enhances perfusion of ischaemic tissues, we subjected mice to femoral artery ligation, a procedure that reduces perfusion of the lower limb, causing ischaemia in the calf (that is, the crural muscle). After ligation, *Phd2*^{+/-} mice showed a milder drop in perfusion and oxygen tension with reduced hypoxia in the crural muscle compared to wild-type (WT) mice (Fig. 1a–g). Ischaemia promotes oxidative stress early and angiogenesis as a later response^{6,11}. Oxidative stress (12 h post-ligation) and capillarization (14 days post-ligation) were both increased in the crural muscle of WT, but not *Phd2*^{+/-} mice (Supplementary Fig. 2a–g). As a consequence of preserved blood flow, *Phd2*^{+/-} crural muscles showed reduced ischaemic necrosis and increased viability (Fig. 1h–j and Supplementary Fig. 2h–j). WT mice showed signs of muscle regeneration that were absent in *Phd2*^{+/-} crural muscles (Supplementary Fig. 2k–m). The protection against ischaemic damage improved physical endurance of *Phd2*^{+/-} mice in ischaemia (Fig. 1k), although both genotypes had similar running capacity at baseline (Supplementary Fig. 2n).

Because *Phd2*^{+/-} mice were protected against ischaemia already 12 h post-ligation, we proposed that they could tolerate ischaemic insults better due to increased collaterals at baseline¹². Macroscopic inspection of the upper limb, that is, the thigh, and histological analysis

of the adductor (in the inner thigh, where collaterals grow) after gelatin-bismuth angiographies showed about twice higher numbers and area of bismuth-positive collaterals in non-ligated *Phd2*^{+/-} versus WT mice (Fig. 1l–q and Supplementary Fig. 3a,b). Also micro-computed tomography scans and X-ray radiographies showed higher numbers of large vessels (>200 μm in diameter) in *Phd2*^{+/-} than WT thighs at baseline (Fig. 1r–t and Supplementary Fig. 3c,d), whereas numbers of smaller vessels (<200 μm in diameter) and capillaries were comparable in both genotypes (Supplementary Fig. 3e–g). Both the total area and numbers of bismuth-positive collaterals were still higher in *Phd2*^{+/-} versus WT adductors 12 and 72 h post-ligation, a time-window when collateral remodelling just begins in WT mice (Fig. 1p,q and Supplementary Fig. 3a,b).

We also assessed whether *Phd2*^{+/-} mice were protected against myocardial ischaemia. Twenty-four hours after coronary artery ligation, desmin-negative area (a readout of cardiomyocyte death) was smaller in *Phd2*^{+/-} hearts (Supplementary Fig. 4a–c). Compared to WT, *Phd2*^{+/-} hearts showed higher perfusion in both infarcted and remote myocardium (Supplementary Fig. 4d–g). At baseline, density of large vessels, but not small vessels and capillaries, was higher in *Phd2*^{+/-} versus WT hearts (Supplementary Fig. 4h–l).

To increase blood flow in case of major arterial occlusion, collateral vessels undergo extensive remodelling (arteriogenesis) with thickening of the tunica media, consisting of α-smooth muscle actin (αSMA)-positive SMCs, and enlargement of vessel diameter⁸. Numbers and total area of αSMA⁺ collateral vessels were higher in *Phd2*^{+/-} adductors both at baseline and after ischaemia, whereas the mean area and thickness of the tunica media were higher only at baseline (Fig. 1u–b'). These data show that collateral vessels of *Phd2*^{+/-} mice at baseline were similar to those of WT mice after femoral artery ligation. This 'collateral vessel preconditioning' was protective against ischaemia.

Inflammatory cells and macrophages in particular are responsible for collateral vessel remodelling^{7,8}. Nevertheless, CD45⁺ leukocyte and F4/80⁺ macrophage infiltration of the adductors was similar at baseline and equally increased after ligation in both genotypes (Fig. 2a,b). We therefore analysed the phenotype of infiltrating macrophages and measured the density of M2-like, wound-healing/pro-angiogenic macrophages by their expression of the MRC1 mannose receptor^{9,10}. At baseline, F4/80⁺MRC1⁺ macrophages were 75% higher in *Phd2*^{+/-} versus WT adductors (Fig. 2c–e). Seventy-two hours after ligation, their numbers were increased by 95% in WT and only 50% in *Phd2*^{+/-} mice

¹Laboratory of Molecular Oncology and Angiogenesis, Vesalius Research Center, VIB, Leuven B-3000, Belgium. ²Laboratory of Molecular Oncology and Angiogenesis, Vesalius Research Center, K. U. Leuven, Leuven B-3000, Belgium. ³Laboratory of Angiogenesis and Neurovascular link, Vesalius Research Center, VIB, Leuven B-3000, Belgium. ⁴Laboratory of Angiogenesis and Neurovascular link, Vesalius Research Center, K. U. Leuven, Leuven B-3000, Belgium. ⁵Life and Health Sciences Research Institute, Minho University, 4710-057 Braga, Portugal. ⁶Atherosclerosis Research Laboratory, CIMA-University of Navarra, 31008 Pamplona, Spain. ⁷Angiogenesis and Tumor Targeting Unit and HSR-TIGET, San Raffaele Institute, 20132 Milan, Italy. ⁸Vita-Salute University, 20132 Milan, Italy. ⁹Biomedical Magnetic Resonance Unit, Université Catholique de Louvain, Brussels B-1200, Belgium. ¹⁰Rayne Institute, University College London, London WC1E 6JF, UK. ¹¹Molecular/Cancer Biology Laboratory, Research Programs Unit and Institute for Molecular Medicine, Biomedicum Helsinki, 00014 Helsinki, Finland. ¹²Cardiovascular Medicine, Yale University School of Medicine, New Haven, Connecticut 06510, USA. ¹³The First Department of Internal Medicine, Nara Medical University, 634-8522 Nara, Japan. †Present addresses: Faculty of Infectious & Tropical Diseases, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK (T.P.); The Swiss Institute for Experimental Cancer Research (ISREC), Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, CH-1015, Switzerland (M.D.P.).

*These authors contributed equally to this work.

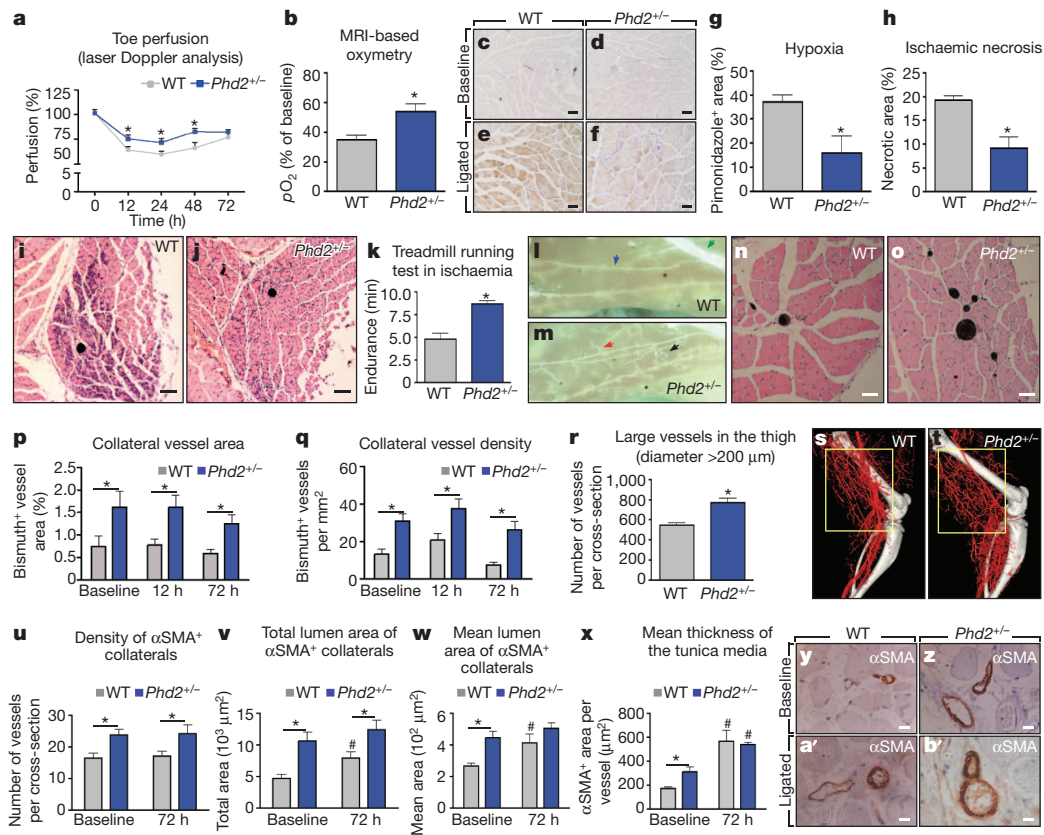


Figure 1 | *Phd2*^{+/-} mice are pre-adapted to ischaemia. **a**, Laser Doppler analysis 12, 24, 48 and 72 h post-ligation. **b**, Magnetic resonance imaging (MRI) oxymetry in crural muscles 12 h post-ligation. **c-g**, Hypoxic area quantification (**g**) on pimonidazole-stained sections of crural muscles 12 h post-ligation (**e, f**); hypoxia was undetectable at baseline (**c, d**). Scale bars, 200 μm. **g**, Pimonidazole⁺ area 12 h post-ligation. **h-j**, Necrotic area quantification (**h**) on haematoxylin- and eosin-stained sections of crural muscles 72 h post-ligation (**i, j**). Scale bars, 100 μm. **k**, Treadmill running test 12 h post-ligation. **l, m**, Macroscopic view of adductors after gelatin-bismuth-based angiographies at baseline. Collateral vessels: primary (blue arrow), secondary (red arrow), and

tertiary (black arrow). Femoral artery (green arrow). **n, o**, Haematoxylin and eosin staining of adductors at baseline after angiographies; bismuth⁺ collaterals appear black. Scale bars, 50 μm. **p, q**, Collateral vessel area (**p**) and density (**q**) represented in **n, o, r**. Quantification of large vessels (>200 μm in diameter) in the thigh at baseline after micro-computed tomography angiograms. **s, t**, Representative micro-computed tomography micrographs of the thigh (yellow frame). **u-x**, Morphological analysis on αSMA-stained sections of adductors at baseline and 72 h post-ligation, as represented in **y-b'**. Scale bars, 10 μm. All graphs show mean ± s.e.m. All experiments, *n* ≥ 5. *, *P* < 0.05 towards WT. #, *P* < 0.05 towards baseline.

(Fig. 2c, f, g). Gene-profiling of *Phd2*^{+/-} peritoneal macrophages (pMΦ) showed higher expression of M2-type genes^{9,10,13}, including *Tek* (also known as *Tie2*), *Arg1*, *Cxcr4*, *Ccr2*, *Hgf*, *Pdgfb*, *Fizz*, *Nrp1*, *Mmp2*, *Cxcl12* (also known as *Sdf1*) and *Tgfb*, than WT pMΦ (Fig. 2h). Conversely, several proinflammatory or anti-angiogenic (that is, M1-type) molecules were downregulated in *Phd2*^{+/-} macrophages; these included *Il1b*, *Il6*, *Nos2* and *Il12* (Fig. 2h). Similarly, *Phd2*^{+/-} macrophages sorted from the adductor expressed higher levels of *Pdgfb*, *Sdf1*, *Tie2*, *Mmp2* and *Nrp1* at baseline (Fig. 2i). Seventy-two hours post-ligation, the expression level of these markers was similar in *Phd2*^{+/-} and WT tissue macrophages (Fig. 2i). Expression of these genes was comparable in WT and *Phd2*^{+/-} endothelial cells isolated from adductors at baseline or in ischaemia (Supplementary Table 1). Noteworthy, the basal level of *Phd2* in *Phd2*^{+/-} macrophages was half of that in WT macrophages and did not change in ischaemia (Fig. 2i). Conversely, *Phd2* expression in WT macrophages was reduced by ~50% in ischaemia and thus reached the same level as in *Phd2*^{+/-} macrophages (Fig. 2i). Thus, *Phd2*^{+/-} macrophages showed a unique and cell-specific gene signature, which was reminiscent, at least in part, of M2-polarized macrophages and of WT macrophages in ischaemia.

We therefore assessed whether WT and *Phd2*^{+/-} macrophages affect the behaviour of endothelial cells and SMCs, the two main cellular components of arteries. Soluble factors released by *Phd2*^{+/-} macrophages strongly increased migration and proliferation of SMCs, but not endothelial cells, probably because the latter were already highly responsive to WT macrophages (Fig. 2j-n and

Supplementary Fig. 5a-e). Consistently, SMCs exposed to conditioned medium from *Phd2*^{+/-} macrophages showed reduced levels of calponin-1, Sm22, smoothelin, NMHC-B and αSMA (Fig. 2o-s), indicating enhanced proliferation^{14,15}. *In vitro* knockdown of both *Sdf1* and *Pdgfb*, known to stimulate SMC recruitment and proliferation^{16,17}, abolished the enhanced response of SMCs to *Phd2*^{+/-} macrophages, although inhibition of either factor was also very effective (Supplementary Fig. 5f, g and Supplementary Note 1). Overall, these data showed that lower levels of PHD2 pre-adapt macrophages to ischaemia by skewing them towards an M2-like phenotype, which promotes SMC recruitment and growth.

We then investigated whether reduced levels of PHD2 in macrophages promoted collateral vessel preconditioning. Myeloid-cell specific *Phd2* haploinsufficiency (*Phd2*^{LysCre;lox/WT}) increased numbers and area of collateral branch arteries, thus conferring protection against ischaemic necrosis and enhancing running capacity in ischaemia (Fig. 3a-i and Supplementary Fig. 6a). In myeloid-cell-specific *Phd2*-null mice (*Phd2*^{LysCre;lox/lox}), arterIALIZATION, ischaemic necrosis and physical endurance were unchanged (Fig. 3a-i and Supplementary Fig. 6a), probably because of the compensatory activity of PHD3, another PHD family member (see below). We also transplanted WT or *Phd2*^{+/-} (hereafter HE for 'heterozygous') bone marrow (BM) cells into irradiated WT (referred to as WT→WT and HE→WT mice, respectively) or *Phd2*^{+/-} (WT→HE and HE→HE mice, respectively) recipients (Supplementary Note 2). Compared to WT→WT, collateral vessel density and area were higher in HE→WT and HE→HE but

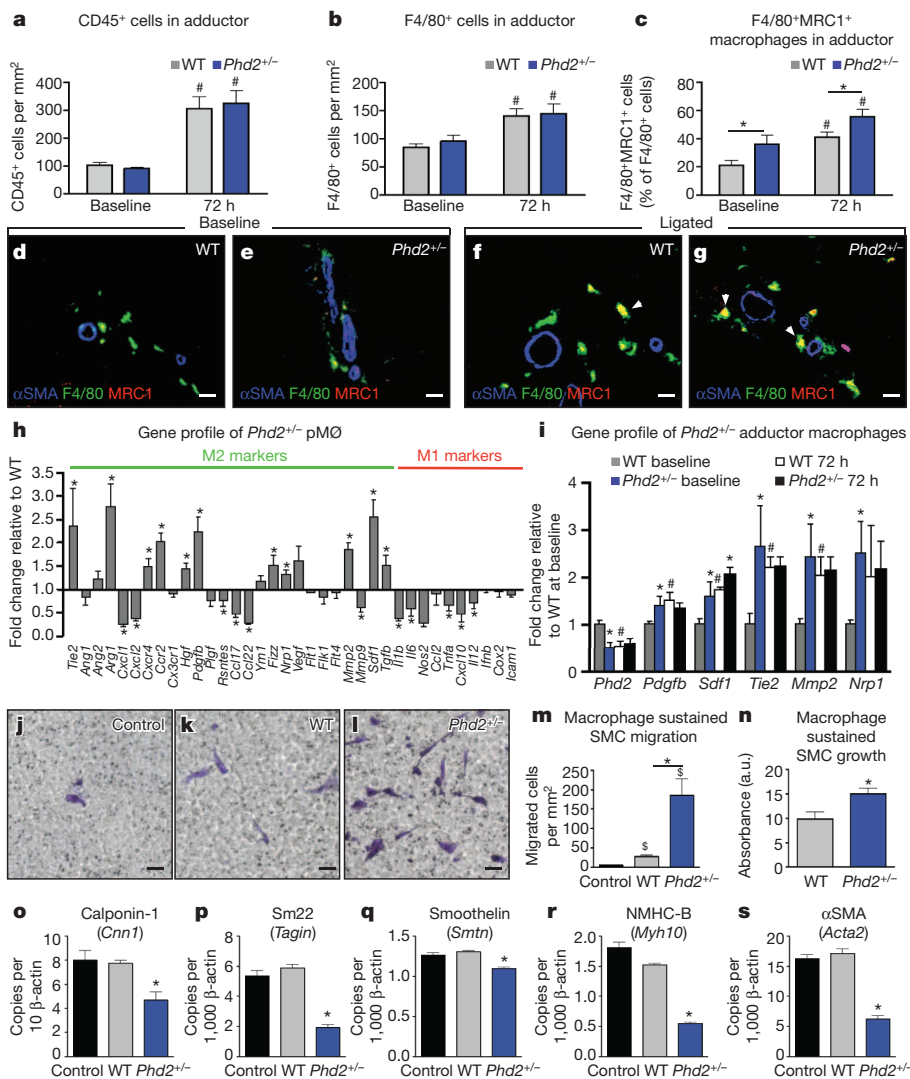


Figure 2 | *Phd2*^{+/-} macrophages display an M2-like phenotype. **a, b**, Quantification of CD45⁺ leukocyte and F4/80⁺ macrophage infiltration of adductors at baseline and 72 h post-ligation. **c**, Quantification of MRC1⁺ macrophages in adductors at baseline and 72 h post-ligation, represented in **d–g**. **d–g**, Micrographs of immunostainings for F4/80 (green), MRC1 (red) and αSMA (blue). Arrowheads (**f, g**) point to F4/80⁺MRC1⁺ cells. Scale bars, 20 μm. **h**, Gene profile by quantitative PCR of peritoneal macrophages (pMφ). **i**, Gene profile of F4/80⁺ macrophages sorted from adductors at baseline and 72 h post-ligation. **j–m**, Quantification (**m**) of crystal-violet-stained SMCs, migrated towards control medium (**j**), WT (**k**) or *Phd2*^{+/-} (**l**) macrophages. Scale bars, 50 μm. **n**, SMC growth in response to soluble factors released by WT or *Phd2*^{+/-} macrophages. **o–s**, Gene profile by quantitative PCR of SMCs exposed to control, WT macrophage-conditioned medium, or *Phd2*^{+/-} macrophage-conditioned medium. All bars show mean ± s.e.m. All experiments, *n* ≥ 5. *, *P* < 0.05 towards WT. #, *P* < 0.05 towards baseline. \$, *P* < 0.05 towards control medium.

similar in WT→HE mice, supporting the key role of BM-derived cells in enhancing collateralization and sustaining pre-existing arteries in *Phd2*^{+/-} mice (Fig. 3j, k). In accordance, HE→HE and HE→WT, but not WT→HE mice, were protected against ischaemic necrosis (Fig. 3l). The running capacity of HE→WT mice in ischaemia was twice as high as in WT→WT mice (Fig. 3m and Supplementary Fig. 6b). Deletion of one *Phd2* allele in haematopoietic cells, endothelial cells or SMCs confirmed that enhanced collateral vessel growth and maintenance were specifically conferred by *Phd2* haploinsufficiency in BM cells, but not endothelial cells or SMCs (Supplementary Tables 2 and 3). We also assessed the effect of acute deletion of *Phd2* in macrophages by transplanting BM cells from tamoxifen-inducible *Phd2*-haploinsufficient mice (*Phd2*^{Rosa26CreERT;lox/WT}) into WT recipient mice (HE^{Rosa26CreERT}→WT). Tamoxifen-induced deletion of one *Phd2* allele in BM cells increased collateral branches and protected against ischaemic necrosis when compared to vehicle (Fig. 3n–p and Supplementary Note 3). Thus, both chronic and acute deletion of one *Phd2* allele in myeloid cells was sufficient to induce pro-arteriogenic macrophages, leading to enhanced collateralization and prevention of ischaemia.

PHD oxygen sensors negatively regulate HIF accumulation and NF-κB activity^{1,11,18–22}. Compared to WT, HIF-1α and HIF-2α protein levels were respectively four and two times higher in *Phd2*-null macrophages, but unchanged in *Phd2*-haploinsufficient macrophages (Fig. 4a). Conversely, NF-κB activity was increased by 65% in *Phd2*-haploinsufficient macrophages but unaffected in *Phd2*-null macrophages (Fig. 4b). We proposed that PHD3 might compensate for the complete loss of *Phd2*

because *Phd3* transcripts were 12.2-fold higher in *Phd2*-null macrophages (Supplementary Fig. 7). Indeed, *Phd3* silencing resulted in a modest induction of NF-κB activity in WT and *Phd2*-haploinsufficient macrophages in contrast to a 70% increase in *Phd2*-null macrophages compared to their scramble controls (Fig. 4b and Supplementary Note 4). PHD2 hydroxylase function was necessary for NF-κB regulation because ectopic expression of a wild-type PHD2 (PHD2^{WT}) greatly blunted the activity of NF-κB luciferase induced by *Phd2* haploinsufficiency, whereas a hydroxylase-deficient PHD2 (PHD2^{H313A}) had no effect (Fig. 4c). NF-κB activation by TNF-α was still significantly stronger in *Phd2*^{+/-} macrophages (Fig. 4d). In contrast, basal and TNF-α-induced NF-κB activity were similar in WT and *Phd2*^{+/-} endothelial cells (Supplementary Fig. 8a). When measuring the nuclear accumulation of NF-κB subunits, we found that members of the canonical pathway, that is, p65 (RelA) and p50 (NF-κB1), were more abundant in *Phd2*^{+/-} than WT macrophages (Fig. 4e). Silencing of p65 or p50 blocked NF-κB hyperactivation in *Phd2*^{+/-} macrophages and the combined knockdown of both subunits restored NF-κB function back to WT levels (Fig. 4f and Supplementary Note 5), thus highlighting the prominent role of NF-κB p65/p50 heterodimers in *Phd2*^{+/-} macrophages. To evaluate the involvement of canonical NF-κB signalling in macrophage skewing by *Phd2* haploinsufficiency, we generated a myeloid-cell specific double transgenic strain, haploinsufficient for *Phd2* and null for *Ikkbb*, the gene encoding IKKβ, a positive regulator of canonical NF-κB pathway. Genetic disruption (or pharmacological inhibition) of canonical NF-κB pathway prevented the upregulation

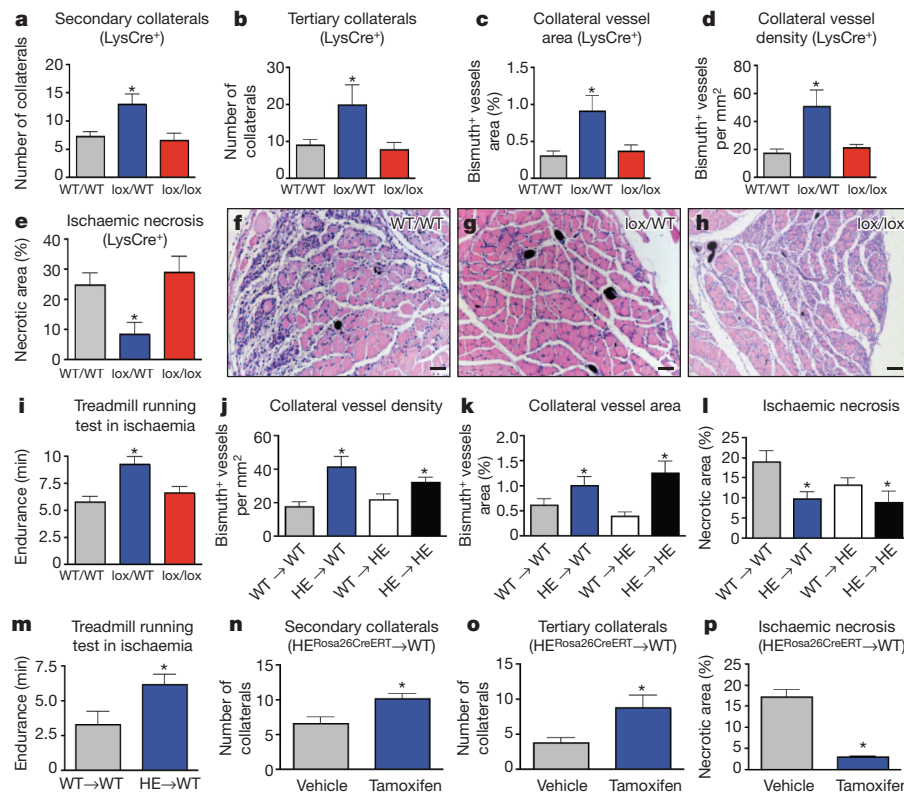


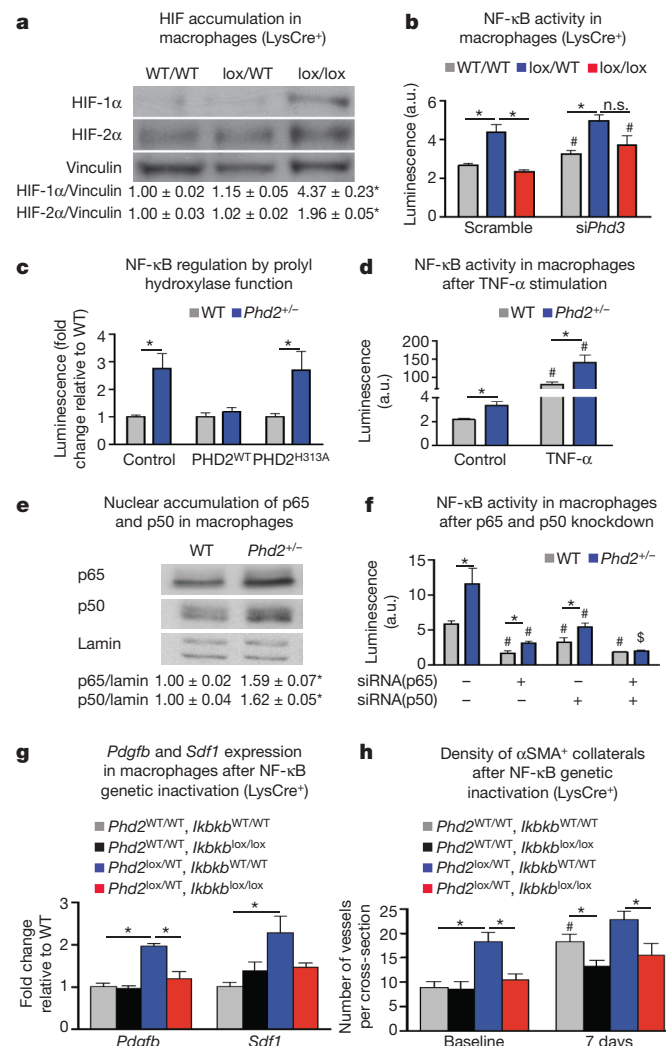
Figure 3 | *Phd2*^{+/-} macrophages protect against ischaemia by inducing arteriogenesis.

a, b, Quantification of secondary (**a**) and tertiary (**b**) collateral branches in WT (*Phd2*^{LysCre;WT/WT}; WT/WT), myeloid-cell specific *Phd2*-haplodeficient (*Phd2*^{LysCre;lox/WT}; lox/WT) or null mice (*Phd2*^{LysCre;lox/lox}; lox/lox). **c, d**, Bismuth⁺ collateral vessel area (**c**) and density (**d**) of adductors at baseline. **e–h**, Necrotic area quantification (**e**) on haematoxylin- and eosin-stained sections of crural muscles 72 h post-ligation (**f–h**). Scale bars, 100 μ m. **i**, Treadmill running test 12 h post-ligation. **j, k**, Bismuth⁺ collateral vessel density (**j**) and area (**k**) of non-occluded limbs 5 weeks after bone marrow (BM) transplantation. *Phd2*^{+/-} BM in WT and *Phd2*^{+/-} recipient mice (HE→WT and HE→HE, respectively); WT BM in WT and *Phd2*^{+/-} recipient mice (WT→WT and WT→HE). **l**, Ischaemic necrosis 72 h post-ligation. **m**, Treadmill running test 12 h post-ligation in WT→WT and HE→WT mice. **n–p**, Quantification of secondary (**n**) and tertiary (**o**) collateral vessels at baseline and of ischaemic necrosis 72 h post-ligation (**p**), following tamoxifen-induced deletion of one *Phd2* allele in BM cells of HE^{Rosa26CreERT}→WT mice. All bars show mean \pm s.e.m. All experiments, $n \geq 4$. *, $P < 0.05$ towards WT/WT and lox/lox in **a–i**, towards WT→WT in **j–m**, or towards vehicle in **n–p**.

of *Pdgfb* and *Sdf1* in cultured *Phd2*-haplodeficient macrophages (Fig. 4g and Supplementary Fig. 8b). *In vivo*, gene inactivation of *Ikbkb* in myeloid cells abolished collateral vessel preconditioning conferred by *Phd2* haplodeficiency and greatly prevented ischaemia-induced arteriogenesis in WT mice (Fig. 4h). Thus, skewing of *Phd2*-haplodeficient macrophages towards a pro-arteriogenic phenotype relied on activation of canonical NF- κ B pathway.

Specific macrophage differentiation states have been implicated in the promotion of angiogenesis during cancer and atherosclerosis progression^{9,10,13,23}. However, little is known of the significance of macrophage heterogeneity in arteriogenesis and its implications on ischaemic diseases. This study identifies a role of macrophage PHD2 in oxygen delivery by regulating arteriogenesis. We show that the phenotype of macrophages induced by reduced PHD2 levels not only favours the formation of new collateral branches, but is also important for collateral vessel homeostasis (Supplementary Note 6). In our model, ischaemia favours the accumulation of M2-like, pro-arteriogenic macrophages that fuel collateral vessel maturation in a NF- κ B-dependent manner (Supplementary Fig. 1). *Phd2* haplodeficiency unleashes constitutive

Figure 4 | *Phd2*^{+/-} macrophages trigger arteriogenesis via canonical NF- κ B pathway. **a**, Western blot for HIF-1 α and HIF-2 α in WT (WT/WT), *Phd2*-haplodeficient (lox/WT) and *Phd2*-null (lox/lox) macrophages. Vinculin was used as loading control. Numbers represent densitometric fold change towards WT/WT. **b**, NF- κ B activity in macrophages after *Phd3* silencing (si*Phd3*; n.s., not significant). **c**, NF- κ B activity in *Phd2*^{+/-} macrophages upon ectopic expression of wild-type PHD2 (PHD2^{WT}) or hydroxylase-deficient PHD2 (PHD2^{H313A}). **d**, NF- κ B activity in macrophages at baseline and upon TNF- α stimulation. **e**, Western blot for nuclear p65 and p50 in WT and *Phd2*^{+/-} macrophages. Numbers represent densitometric fold change towards WT. **f**, NF- κ B activity in macrophages after silencing of p65, p50, or combination of both. **g**, *Pdgfb* and *Sdf1* expression (quantitative PCR) after genetic inactivation of *Ikbkb* in *Phd2*-haplodeficient pM ϕ (*Phd2*^{lox/WT}; *Ikbkb*^{lox/lox}). **h**, Quantification of α SMA⁺ collaterals in myeloid-cell-specific *Phd2*-haplodeficient and *Ikbkb*-null mice at baseline and 7 days post-ligation. All bars and values show mean \pm s.e.m. All experiments, $n \geq 4$. *, $P < 0.05$ towards wt/wt in **a, b, g, h**, or towards WT in **c–f**. #, $P < 0.05$ towards scramble in **b, f**, or towards baseline in **d, h**. \$, $P < 0.05$ towards scramble and either short interfering RNA alone.



NF- κ B signals that pre-adapt tissue-resident macrophages to ischaemia, accounting for the enhanced arteriogenesis at baseline and thus protection against ischaemic tissue demise. In particular, we show that NF- κ B activation in *Phd2*^{+/-} macrophages increases the production of SDF1 and PDGFB, which are concurrently required to prime SMC migration and growth *in vitro*. The relevance of SDF1 and PDGFB *in vivo* remains to be established, and other soluble factors may also contribute.

Repression of canonical NF- κ B pathway in macrophages promotes upregulation of M1-type and downregulation of M2-type genes^{24–26}. PHDs negatively regulate NF- κ B through either hydroxylase-dependent or -independent inactivation of IKK β in different cell contexts^{19–22}. We show here that *Phd2* haploinsufficiency results in hyperactivation of canonical NF- κ B pathway in macrophages and promotion of the M2-phenotype via accumulation of both p65 and p50 subunits, and that this regulation requires PHD hydroxylase function.

In summary, our study provides an insight on how the PHD2 oxygen sensor regulates arteriogenesis by modulating macrophage phenotype. The mechanism upstream to arteriogenic PHD2 downregulation in M2-like macrophages remains to be established. Nonetheless, our findings support the rationale for therapeutic inhibition of PHD2. Previous studies showed that unspecific inhibitors of PHD2 or silencing of PHD2 promote therapeutic revascularization against ischaemia^{2–5}. However, angiogenesis is a late response and organ function might be compromised before new blood vessel formation is achieved. In contrast, arteriogenesis takes place on pre-existing vascular shunts and our data suggest that either PHD2 inhibitors or cell-therapy-based strategies using PHD2 hypomorphic macrophages might be exploited as preventive medicine to promote collateral vascularization in patients at risk of limb or heart ischaemia, such as diabetic or hypercholesterolemic patients.

METHODS SUMMARY

129/S6 or Balb/c WT and *Phd2*^{+/-} mice (8–12 weeks old) were obtained from our mouse facility. *Phd2*^{+/-} and *Phd2* conditional knockout mice were obtained as previously described¹. To induce hindlimb ischaemia, unilateral or bilateral ligations of the femoral artery and vein and the cutaneous vessels branching from the caudal femoral artery side branch were performed without damaging the nervus femoralis²⁷. Oxygen tension (pO₂) in the lower limb was measured 12 h after femoral artery ligation by ¹⁹F magnetic resonance imaging oxymetry. Adductors and crural muscles were dissected, fixed in 2% paraformaldehyde, dehydrated, embedded in paraffin and sectioned at 7- μ m thickness for histology, immunostaining and morphometry analysis. Macrophages were either collected from the peritoneal cavity of the mice (pMO) or derived from BM precursors as described before²⁸. Balb/c WT and *Phd2*^{+/-} recipient mice were irradiated with 7.5 Gy. Subsequently, 5 \times 10⁶ BM cells from green fluorescent protein⁺ (GFP⁺) WT or GFP⁺ *Phd2*^{+/-} mice were injected intravenously via the tail vein. Femoral artery ligation, treadmill running test and bismuth angiography were performed 5 weeks after BM reconstitution. Full Methods and any associated references are available in the Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 September 2010; accepted 23 August 2011.

Published online 9 October 2011.

- Mazzone, M. *et al.* Heterozygous deficiency of PHD2 restores tumor oxygenation and inhibits metastasis via endothelial normalization. *Cell* **136**, 839–851 (2009).
- Milkiewicz, M., Pugh, C. W. & Egginton, S. Inhibition of endogenous HIF inactivation induces angiogenesis in ischaemic skeletal muscles of mice. *J. Physiol. (Lond.)* **560**, 21–26 (2004).
- Nangaku, M. *et al.* A novel class of prolyl hydroxylase inhibitors induces angiogenesis and exerts organ protection against ischemia. *Arterioscler. Thromb. Vasc. Biol.* **27**, 2548–2554 (2007).
- Huang, M. *et al.* Short hairpin RNA interference therapy for ischemic heart disease. *Circulation* **118**, S226–S233 (2008).
- Loinard, C. *et al.* Inhibition of prolyl hydroxylase domain proteins promotes therapeutic revascularization. *Circulation* **120**, 50–59 (2009).
- Carmeliet, P. Mechanisms of angiogenesis and arteriogenesis. *Nature Med.* **6**, 389–395 (2000).
- Simons, M. & Ware, J. A. Therapeutic angiogenesis in cardiovascular disease. *Nature Rev. Drug Discov.* **2**, 863–872 (2003).

- Schaper, W. Collateral circulation: past and present. *Basic Res. Cardiol.* **104**, 5–21 (2009).
- Mantovani, A. & Sica, A. Macrophages, innate immunity and cancer: balance, tolerance, and diversity. *Curr. Opin. Immunol.* **22**, 231–237 (2010).
- Nucera, S., Biziato, D. & De Palma, M. The interplay between macrophages and angiogenesis in development, tissue injury and regeneration. *Int. J. Dev. Biol.* **55**, 495–503 (2011).
- Aragónés, J. *et al.* Deficiency or inhibition of oxygen sensor Phd1 induces hypoxia tolerance by reprogramming basal metabolism. *Nature Genet.* **40**, 170–180 (2008).
- Helisch, A. *et al.* Impact of mouse strain differences in innate hindlimb collateral vasculature. *Arterioscler. Thromb. Vasc. Biol.* **26**, 520–526 (2006).
- Squadrito, M. L. & De Palma, M. Macrophage regulation of tumor angiogenesis: Implications for cancer therapy. *Mol. Aspects Med.* **32**, 123–145 (2011).
- Kumar, M. S. & Owens, G. K. Combinatorial control of smooth muscle-specific gene expression. *Arterioscler. Thromb. Vasc. Biol.* **23**, 737–747 (2003).
- Wolf, C. *et al.* Vascular remodeling and altered protein expression during growth of coronary collateral arteries. *J. Mol. Cell. Cardiol.* **30**, 2291–2305 (1998).
- Karshovska, E., Zagorac, D., Zerneck, A., Weber, C. & Schober, A. A small molecule CXCR4 antagonist inhibits neointima formation and smooth muscle progenitor cell mobilization after arterial injury. *J. Thromb. Haemost.* **6**, 1812–1815 (2008).
- Hellstrom, M., Kalen, M., Lindahl, P., Abramsson, A. & Betsholtz, C. Role of PDGF-B and PDGFR- β in recruitment of vascular smooth muscle cells and pericytes during embryonic blood vessel formation in the mouse. *Development* **126**, 3047–3055 (1999).
- Jaakkola, P. *et al.* Targeting of HIF- α to the von Hippel-Lindau ubiquitylation complex by O₂-regulated prolyl hydroxylation. *Science* **292**, 468–472 (2001).
- Chan, D. A. *et al.* Tumor vasculature is regulated by PHD2-mediated angiogenesis and bone marrow-derived cell recruitment. *Cancer Cell* **15**, 527–538 (2009).
- Xue, J. *et al.* Prolyl hydroxylase-3 is down-regulated in colorectal cancer cells and inhibits IKK β independent of hydroxylase activity. *Gastroenterology* **138**, 606–615 (2010).
- Cummins, E. P. *et al.* Prolyl hydroxylase-1 negatively regulates I κ B kinase- β , giving insight into hypoxia-induced NF κ B activity. *Proc. Natl Acad. Sci. USA* **103**, 18154–18159 (2006).
- Fu, J. & Taubman, M. B. Prolyl hydroxylase EGLN3 regulates skeletal myoblast differentiation through an NF- κ B-dependent pathway. *J. Biol. Chem.* **285**, 8927–8935 (2010).
- Mantovani, A., Garlanda, C. & Locati, M. Macrophage diversity and polarization in atherosclerosis: a question of balance. *Arterioscler. Thromb. Vasc. Biol.* **29**, 1419–1423 (2009).
- Fong, C. H. *et al.* An antiinflammatory role for IKK β through the inhibition of “classical” macrophage activation. *J. Exp. Med.* **205**, 1269–1276 (2008).
- Hagemann, T. *et al.* “Re-educating” tumor-associated macrophages by targeting NF- κ B. *J. Exp. Med.* **205**, 1261–1268 (2008).
- Greten, F. R. *et al.* NF- κ B is a negative regulator of IL-1 β secretion as revealed by genetic and pharmacological inhibition of IKK β . *Cell* **130**, 918–931 (2007).
- Luttun, A. *et al.* Revascularization of ischemic tissues by PIGF treatment, and inhibition of tumor angiogenesis, arthritis and atherosclerosis by anti-Fit1. *Nature Med.* **8**, 831–840 (2002).
- Meerpohl, H. G., Lohmann-Matthes, M. L. & Fischer, H. Studies on the activation of mouse bone marrow-derived macrophages by the macrophage cytotoxicity factor (MCF). *Eur. J. Immunol.* **6**, 213–217 (1976).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from FWO (G.0726.10), Belgium, and from VIB. The authors are thankful to P. Carmeliet for the *Phd2* KO and cKO mice, M. Karin for the *Ikbkb* cKO mice, P. Ratcliffe for the PHD2^{H313A} construct, A. Luttun and P. Fazzari for comments, Y. Jonsson and T. Janssens for technical assistance. VE-Cadherin:CreERT and PDGFRB:Cre transgenic mice were generated at Cancer Research UK and kindly donated by R. Adams. E.D. was funded by ARC, S.C. by FCT, R.L.O. and V.F. by FWO, A.H. by DFG. C.R. was supported by COST action TD0901. M.D.P. was supported by an ERC starting grant.

Author Contributions Y.T., E.D. and S.C. performed experimental design, all experiments, acquisition of data and analysis and interpretation of all data. C.R. performed analysis of histological stainings, angiographies. R.L.O., C.R. and S.C. performed the western blots. R.L.O. and V.F. performed treadmill-running tests, quantitative PCR experiments and drug administrations. M.L.S. performed lentiviral vector preparation and cell transduction. F.B. performed EC isolation and angiography measurements. J.M., B.F.J. and B.G. performed oxymetry experiments. S.D. performed luciferase assays. M.W. and A.H. performed transplantation experiments and electroporations. Y.T. and J.S. performed the ligations of the femoral artery. Z.W.Z. and M.S. performed micro-computer tomography angiograms. A.A. and K.A. contributed vital reagents. T.B. and P.M. contributed in generating the *Phd2* targeting vector. Y.T., E.D., S.C., C.R., Y.S. and M.D.P. participated in scientific discussion and drafting of the manuscript. M.M. performed experimental design, data analysis, conducted scientific direction and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.M. (massimiliano.mazzone@vib-kuleuven.be)

METHODS

Animals. 129/S6 or Balb/c, female and male, WT and *Phd2*^{+/-} mice (8–12 weeks old) were obtained from our mouse facility. *Phd2*^{+/-} and *Phd2* conditional knockout mice were obtained as previously described¹. VE-Cadherin:CreERT and PDGFRB:Cre transgenic mice were generated by R. Adams at the Cancer Research UK^{29,30}. *Ikbkb* conditional knockout mice were obtained from M. Karin³¹. Tie2:Cre and Rosa26:CreERT transgenic mice were purchased from the Jackson Laboratory. Housing and all experimental animal procedures were performed in accordance with Belgian law on animal care and were approved by the Institutional Animal Care and Research Advisory Committee of the K. U. Leuven (P036/2009).

Mouse model of hindlimb ischaemia and myocardial infarction. To induce hindlimb ischaemia, unilateral or bilateral ligations of the femoral artery and vein (proximal to the popliteal artery) and the cutaneous vessels branching from the caudal femoral artery side branch were performed without damaging the nervus femoralis. By using this procedure, collateral flow to adductor muscles is preserved via arterioles branching from the femoral artery, therefore 50% up to 60% of perfusion is preserved by this method. Two superficial pre-existing collateral arterioles, connecting the femoral and saphenous artery, were used for analysis. Functional perfusion measurements of the collateral region were performed using a Lisca PIM II camera (Gambro). Gelatin-bismuth-based angiography was performed as described before and analysed by photoangiographs (Nikon D1 digital camera). Collateral side branches were categorized as follows: secondary collateral arterioles directly branched from the primary collateral, and tertiary arterioles orientated perpendicularly to the secondary branches. The number of secondary and tertiary collateral arterioles was counted. After perfusion-fixation, the muscle tissue between the two superficial collateral arterioles (adductor) was post-fixed in 2% paraformaldehyde, paraffin-embedded and morphometrically analysed²⁷. An endurance treadmill-running test was performed at baseline and 12 h post-bilateral-ligation. Myocardial infarction was induced by permanent ligation of the left anterior descending coronary artery as previously described³². Briefly, the left thorax of anesthetized mice was opened in the fourth intercostal space and all the muscles overlying the intercostal region were dissected. The main left anterior descending coronary artery was ligated proximal to the main bifurcation through a small incision of the pericardium. Discoloration of the ventricle after blood flow restriction was used as readout of a successful surgical procedure. Gelatin-bismuth-based angiography was performed 24 h post-ligation and the entire heart was fixed in 2% paraformaldehyde.

Oxymetry. Oxygen tension (pO₂) in the lower limb was measured using ¹⁹F-MRI oxymetry in non-ligated and ligated legs 12 h after femoral artery ligation. The oxygen reporter probe hexafluorobenzene was injected directly into the crural muscle. Magnetic resonance imaging was performed with a 4.7 T (200 MHz, ¹H), 40-cm inner diameter bore system (Bruker Biospec). A tunable ¹H/¹⁹F surface coil was used for radiofrequency transmission and reception³³.

Histology, immunostaining, and morphometry. Adductor, crural muscles and hearts were dissected, fixed in 2% paraformaldehyde, dehydrated, embedded in paraffin, and sectioned at 7-μm thickness. Area of necrotic tissues in the crural muscle was analysed by haematoxylin and eosin staining. Necrotic cells display a more glassy homogeneous appearance in the cytoplasm with increased eosinophilia, whereas the nuclear changes are reflected by karyolysis, pyknosis and karyorrhexis. Necrotic area was defined as the percentage of area which includes these necrotic myocytes, inflammatory cells and interstitial cells, compared to the total soleus area. Infarct size was measured in desmin-stained hearts 24 h after ischaemia as previously described³⁴. After deparaffinization and rehydration, sections were blocked and incubated overnight with primary antibodies: rat anti-CD31, dilution 1/500 (BD-pharmingen); mouse anti-αSMA, dilution 1/500 (Dako); rat anti-F4/80, dilution 1/100 (Serotec); dilution 1/50 (BD-pharmingen); rat anti-CD45, dilution 1/100 (BD-pharmingen); goat anti-MRC1, dilution 1/200 (R&D); rabbit anti-desmin dilution 1/150 (Cappel). To analyse capillary density and area, images of anti-CD31-stained sections of the entire soleus were taken at ×40. To measure bismuth-positive vessel density and area, haematoxylin- and eosin-stained paraffin sections were analysed and vessels filled with bismuth-gelatin (black spots) were taken in account. Images of the entire soleus were acquired at ×20 for this analysis. The values in the graph represent the averages of the mean vessel density and area per soleus muscle. The same method was used to quantify vessel capillaries and collateral branches in cardiac tissues. Density and area were measured by using a KS300 (Leica) software analysis. Hypoxic cells were analysed 2 h after injection of 60 mg kg⁻¹ pimonidazole into operated mice. Mice were killed and muscles were collected. Paraffin sections were stained with Hypoxiprobe-1-Mab-1 (Hypoxiprobe kit; Chemicon International) following the manufacturer's instructions. Oxidative stress and proliferation rate were assessed on 7-μm thick cryosections by using the goat anti-8-OHdG antibody, dilution 1/200 (Serotec) and the rat anti-BrdU antibody, dilution 1/300 (Serotec). Sections were subsequently incubated with appropriate secondary

antibodies, developed with fluorescent dyes or 3,3'-disminobenzidine (DAB, Sigma). Whole-muscle viability was assessed on unfixed 2-mm thick tissue slices by staining with 2,3,5-triphenyltetrazolium chloride (TTC). Viable area, stained in red, was traced and analysed. Pictures for morphometric analysis were taken using a Retiga EXi camera (Q Imaging) connected to a Nikon E800 microscope or a Zeiss Axio Imager connected to an AxioCam MRc5 camera (Zeiss) and analysis was performed using KS300 (Leica). Angiograms were obtained by X-ray and micro-computed tomography angiographies of hearts and legs at baseline.

Macrophage preparation. To harvest peritoneal macrophages (pMØ), the peritoneal cavity was washed with 5 ml of RPMI 10% FBS. The pooled cells were then seeded in RPMI 10% FBS in 6-well plates (2 × 10⁶ cells per well), 12-well plates (1 × 10⁶ cells per well), or 24-well plates (5 × 10⁵ cells per well). After 6 h of incubation at 37 °C in a moist atmosphere of 5% CO₂ and 95% air, non-adhering cells on each plate were removed by rinsing with phosphate-buffered saline (PBS). The attached macrophages were grown in different media for 12 h or 48 h depending on the experiments performed, as described below. When high amounts of cells were needed (analysis for HIF accumulation and NF-κB activity), macrophages were derived from bone marrow (BM) precursors as described before²⁸. Briefly, BM cells (2 × 10⁶ cells per ml) were grown in a volume of 5 ml in a 10-cm Petri dish (non tissue culture treated, bacterial grade) for 7 to 10 days in DMEM supplemented with 20% FBS and 30% L929 conditioned medium as a source of M-CSF. The cells obtained in those cultures are uniformly macrophages.

Quantitative PCR analysis. In order to investigate gene expression in pMØ, quantitative real-time PCR (qPCR) was performed. After preparing pMØ, the cells were cultured in normoxic condition for 12 h and RNA was extracted. To analyse the expression levels of chemoattractants in the adductor, tissues were collected at baseline or 24 h post-ischaemia and RNA was extracted. Macrophages and endothelial cells were freshly sorted from dissected adductors as described below and RNA was extracted. Quantitative PCR was performed with commercially available or home-made primers and probes for the studied genes. The assay ID (Applied Biosystems) or the sequence of primers and probes (when custom-made) are listed in Supplementary Table 4. RNA levels of *Sdf1* and *Pdgfrb* after inhibition of NF-κB pathway were measured by qPCR on pMØ exposed for 12 h to 500 nM 6-amino-4-(4-phenoxyphenylethylamino)quinazoline.

Protein extraction and immunoblot. Protein extraction was performed using 8 M urea buffer (10% glycerol, 1% SDS, 5 mM DTT, 10 mM Tris-HCl, pH 6.8) as previously described¹. Nuclear proteins were extracted in 1% SDS buffer upon cytoplasmic separation by using a hypotonic lysis buffer (10 mM KCl, 10 mM EDTA, 0.5% NP40, 10 mM HEPES, pH 8, supplemented with phosphatase and protease inhibitors, from Roche). Signal was detected using the ECL system (Invitrogen) according to the manufacturer's instructions. The following antibodies were used: rabbit anti-HIF-1α (Novus), rabbit anti-p105/50, rabbit anti-HIF-2α (Abcam), PM9 rabbit anti-HIF-2α (from P.M.), mouse anti-vinculin (Sigma), rabbit anti-lamin A/C, rabbit anti-p65 (Cell Signaling). Densitometric analysis was performed using ImageJ 1.44 (<http://rsbweb.nih.gov/ij/>).

Transduction and transfection of BM-derived macrophages and lung endothelial cells. To express an inducible NF-κB-responsive firefly luciferase reporter, commercially available lentiviral vectors (LV) were used (Cignal Lenti NF-κB Reporter; SABiosciences). BM-derived macrophages (2.5 × 10⁵) and 10⁵ primary lung endothelial cells, isolated as described before¹, were seeded in a 24-well plate in DMEM 10% FBS or M199 20% FBS for 8 h. Cells were transduced with 10⁸ transducing units per ml. Eight hours after transduction, the medium was replaced. After 48 h, cells were stimulated with TNF-α (20 ng ml⁻¹) for 8 h and the same amount of protein extract was read in a luminometer. For PHD3 silencing, siRNA oligonucleotides were designed using the Invitrogen online siRNA design tool (<http://rnaidesigner.invitrogen.com>). The following siRNA sequences (sense strands) were used. For *Egln3/Phd3* (NM_028133.2): 5'-GCCGGCTGGGCAAACTACTATGCTA-3'; for the scramble: 5'-CACCGCTTAACCCGATTGCGCTAT-3'. In brief, one day after the transduction of macrophages with LV, cells were transfected using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Preparation of the oligonucleotide-Lipofectamine 2000 complexes was done as follows: 25 pmol siRNA oligonucleotide (stock 20 μM) was diluted in 50 μl Opti-MEM I reduced serum medium. Lipofectamine 2000 (1.5 μl) was diluted in 50 μl Opti-MEM I reduced serum medium and incubated for 5 min at room temperature. siRNA oligonucleotides were gently mixed with Lipofectamine 2000 and allowed to incubate at room temperature for 20 min to form complexes. Just before transfection, the cell culture medium was removed and cells were rinsed twice with serum-free Opti-MEM I medium. The Lipofectamine 2000-siRNA oligonucleotide complexes were added to each well in 400 μl of serum-free Opti-MEM medium for 5 h. Afterwards, cells were incubated in complete medium for 48 h at 37 °C in a CO₂ incubator and assayed for gene knockdown (qPCR) and luciferase activity. To assess if the increased NF-κB activity observed in *Phd2*^{+/-} macrophages was dependent on the hydroxylase activity of PHD2, 48 h before transduction,

4×10^6 BM-derived macrophages were resuspended in 240 μ l of Opti-MEM and were electroporated (250 V, 950 μ F, $\infty \Omega$) with 7 μ g of plasmids expressing a wild-type PHD2 (PHD2^{WT}) or a PHD2 containing a mutation at the catalytic site (PHD2^{H313A}). Silencing of the canonical pathway subunits p65 (*Rela*) and p50 (*Nfkb1*) was achieved by electroporation with specific siRNAs. Briefly, 48 h before transduction, 2.4×10^6 BM-derived macrophages were resuspended in 320 μ l of Opti-MEM and were electroporated (250 V, 950 μ F, $\infty \Omega$) with 60 pmol of siRNA for either scramble, p65, p50, or combination of p50 and p65. For higher efficiency of silencing, two different siRNA sequences for each respective gene were designed (<http://rnaidesigner.invitrogen.com>). For p65 (NM_009045.4): 5'-TGTCTGCACC TGTTCCAAATT-3' and 5'-TGCTGATGGAGTACCCTGATT-3'; for p50 (NM_008689.2): 5'-GAATACTTCATGTGACTAATT-3' and 5'-CAAAGGTTAT CGTTCAGTTTT-3'; for the scramble: 5'-CACCGCTTAACCCGTATTGCCTAT-3'.

Cell migration and viability assays. Migration and proliferation of smooth muscle cells (SMCs) and endothelial cells were assessed by using 8- μ m-pore Transwell permeable plate for migration assays and 0.4- μ m-pore Transwell permeable plate for proliferation assays (Corning Life Science). To determine cell migration and proliferation in response to soluble factors secreted by pMØ, pMØ were pre-cultured in the lower chamber for 12 h in RPMI 1% FBS or in M-199 1% FBS (migration assay), or 48 h in DMEM-F12 1% FBS or in M-199 1% FBS (proliferation assay). For migration assays, hCASMCs (human coronary artery SMCs; from Lonza) and HUVECs (human umbilical vein endothelial cells; from Lonza) were starved for 12 h in their own medium at 1% FBS and then seeded in the upper chamber (5×10^3 cells in 200 μ l of medium 1% FBS). SMCs and HUVECs were incubated for 2 days or 24 h, respectively, and migrated cells were fixed with 4% paraformaldehyde, stained with 0.25% crystal violet/50% methanol and counted under the microscope. For cell growth assays, RAOSMCs (rat aortic SMCs) and HUVECs were seeded on the upper chambers (5×10^3 cells per transwell) and cultivated with pMØ for 24 h in DMEM-F12 1% FBS or M-199 1% FBS for RAOSMCs and HUVECs, respectively. The cell proliferative ability was then analysed using WST-1 Cell Proliferation Assay (Roche Applied Biosciences) according to the manufacturer's instructions. Alternatively, WT and *Phd2*^{+/-} pMØ were seeded in the lower chamber of a Transwell and transduced with lentiviral vectors (10^8 TU/ml; Sigma) carrying an shRNA against *Sdf1* (NM_013655.4), *Pdgfr* (NM_011057.3), or a scramble control (purchased from Sigma; TRCN0000178772 for *Sdf1*: 5'-CCGGCTGAAGAACAACAACAGACAA CTCGAGTTGCTGTGTGTTGTTCTTCAGTTTTTTG-3'; TRCN0000042529 for *Pdgfr*: 5'-CCGGGAGTCGAGTTGGAAGCTCATCTCGAGATGAGCT TTCCAACCTCGACTCTTTTTG-3'; SHC002V for the scramble: 5'-CCGGCA ACAAGATGAAGAGCACCACCTCGAGTTGGTGCTCTTCATCTTGTGTT TTT-3'). Sixty hours after macrophage transduction, SMC migration or growth assays were performed by seeding the SMCs in the upper side of the Transwell as above.

SMC differentiation assay. pMØ were seeded in a 24-well plate with DMEM F-12 5% FBS. Conditioned medium was collected after 2 days and supplemented with

25 mM HEPES. RAOSMCs were seeded in a 24-well plate (80×10^3 cells per well) and incubated for 5 h at 37 °C in a moist atmosphere of 5% CO₂ and 95% air. After 2 h of starvation in DMEM-F12 1% FBS, SMCs were stimulated with conditioned medium from WT and *Phd2*^{+/-} pMØ. After 24 h, differentiation status of the SMCs was assessed by qPCR.

Macrophage and endothelial cell sorting. For cell sorting of adductor macrophages and endothelial cells, the adductors were dissected, dissociated mechanically, and digested using collagenase I for 45 min at 37 °C. For macrophage sorting, the digested cell suspension was incubated for 15 min with mouse anti-CD16/CD32 mAb (Fc Block, BD-pharmingen) and stained with rat FITC-conjugated anti-F4/80 antibody (Serotec) for 20 min at 4 °C. CD31⁺CD45⁻ endothelial cells were sorted from the digested adductor cell suspension after incubation with rat APC-conjugated anti-CD31 and rat FITC-conjugated anti-CD45 (BD-pharmingen) for 20 min at 4 °C.

BM transplantation and haematological analysis. Balb/c WT and *Phd2*^{+/-} recipient mice were irradiated with 7.5 Gy. Subsequently, 5×10^6 BM cells from green fluorescent protein⁺ (GFP⁺) WT or GFP⁺ *Phd2*^{+/-} mice were injected intravenously via the tail vein. Femoral artery ligation, treadmill running test and bismuth angiography were performed 5 weeks after BM reconstitution. Red and white blood cell count was determined using a haemocytometer (Cell-Dyn 3700, Abbott) on peripheral blood collected in heparin by retro-orbital bleeding. To assess the effect on arteriogenesis of acute deletion of one *Phd2* allele in macrophages, 5×10^6 BM cells from *Phd2*^{Rosa26CreERT²lox/WT} mice were transplanted into lethally irradiated WT recipient mice. After 5 weeks, transplanted mice were injected intraperitoneally with tamoxifen (1 mg per mouse; Sigma) or vehicle for 5 days. Femoral artery ligation was performed 10 days after tamoxifen treatment as above.

Statistics. The data were represented as mean \pm s.e.m. of the indicated number of measurements. Statistical significance was calculated by two-tailed unpaired *t*-test for two data sets and ANOVA followed by Bonferroni post hoc test for multiple data sets using Prism (GraphPad Inc.), with *P* < 0.05 considered statistically significant.

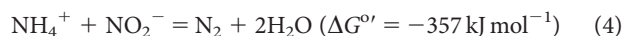
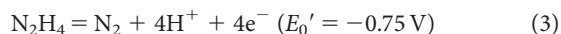
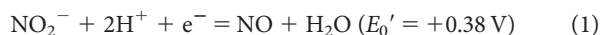
29. Foo, S. S. *et al.* Ephrin-B2 controls cell motility and adhesion during blood-vessel wall assembly. *Cell* **124**, 161–173 (2006).
30. Bénédicto, R. *et al.* The notch ligands Dll4 and Jagged1 have opposing effects on angiogenesis. *Cell* **137**, 1124–1135 (2009).
31. Chen, L. W. *et al.* The two faces of IKK and NF- κ B inhibition: prevention of systemic inflammation but increased local injury following intestinal ischemia-reperfusion. *Nature Med.* **9**, 575–581 (2003).
32. Lutgens, E. *et al.* Chronic myocardial infarction in the mouse: cardiac structural and functional changes. *Cardiovasc. Res.* **41**, 586–593 (1999).
33. Jordan, B. F., Cron, G. O. & Gallez, B. Rapid monitoring of oxygenation by 19F magnetic resonance imaging: Simultaneous comparison with fluorescence quenching. *Magn. Reson. Med.* **61**, 634–638 (2009).
34. Pfeffer, M. A. *et al.* Myocardial infarct size and ventricular function in rats. *Circ. Res.* **44**, 503–512 (1979).

Molecular mechanism of anaerobic ammonium oxidation

Boran Kartal¹, Wouter J. Maalcke¹, Naomi M. de Almeida¹, Irina Cirpus¹, Jolein Gloerich², Wim Geerts¹, Huub J. M. Op den Camp¹, Harry R. Harhangi¹, Eva M. Janssen-Megens³, Kees-Jan Francoijs³, Hendrik G. Stunnenberg³, Jan T. Keltjens¹, Mike S. M. Jetten^{1,4} & Marc Strous^{1,5}

Two distinct microbial processes, denitrification and anaerobic ammonium oxidation (anammox), are responsible for the release of fixed nitrogen as dinitrogen gas (N₂) to the atmosphere^{1–4}. Denitrification has been studied for over 100 years and its intermediates and enzymes are well known⁵. Even though anammox is a key biogeochemical process of equal importance, its molecular mechanism is unknown, but it was proposed to proceed through hydrazine (N₂H₄)^{6,7}. Here we show that N₂H₄ is produced from the anammox substrates ammonium and nitrite and that nitric oxide (NO) is the direct precursor of N₂H₄. We resolved the genes and proteins central to anammox metabolism and purified the key enzymes that catalyse N₂H₄ synthesis and its oxidation to N₂. These results present a new biochemical reaction forging an N–N bond and fill a lacuna in our understanding of the biochemical synthesis of the N₂ in the atmosphere. Furthermore, they reinforce the role of nitric oxide in the evolution of the nitrogen cycle.

Ammonium is difficult to activate in the absence of molecular oxygen. Therefore, how anammox bacteria are able to oxidize ammonium coupled to the reduction of nitrite and forge an N–N bond to make N₂ has been an intriguing question for a long time. Based on the *in silico* analysis of the genome assembly of the anammox bacterium *Kuenenia stuttgartiensis*, a set of three redox reactions (equations (1)–(3)) involving N₂H₄ and nitric oxide (NO) was proposed⁶ to explain the overall anammox stoichiometry (equation (4)):



The role of N₂H₄ in anammox catabolism was originally proposed based on the observation that the compound transiently accumulated when anammox bacteria were incubated with millimolar quantities of hydroxylamine^{7,8}. However, the turnover of neither N₂H₄, hydroxylamine nor NO was demonstrated to start from the actual substrates ammonium and nitrite; thus it remained unclear whether the observed reaction was an integral part of the anammox pathway or a side reaction.

In the present study, we resolved the anammox pathway and its enzymes by a combination of complementary approaches (Fig. 1). *K. stuttgartiensis* was enriched and grown as suspended cells in a membrane bioreactor^{9,10}. Fluorescence *in situ* hybridization (FISH) showed that *K. stuttgartiensis* made up more than 95% of the population. Transcription was shown for more than 97% of all genes after random hexamer-primed reverse transcription of extracted RNA, sequencing and mapping of 5.6 million 32-nucleotide reads on an Illumina Genome Analyser (metatranscriptome accession number

GSE15408). Expression of 1010 proteins was demonstrated by meta-proteomics¹¹ (peptidome accession number PSE111). Further, inhibitor and isotope labelling studies were performed and the activity of enzyme complexes was demonstrated after their purification by liquid chromatography.

Transcriptomics and proteomics indicated that *K. stuttgartiensis* expressed *cd1* nitrite:nitric oxide reductase (NirS, kuste4136, 9% of predicted peptides detected (p.p.d.) and 6.3-fold messenger RNA (mRNA) coverage) with the potential ability to reduce nitrite to NO. This possibility was investigated by incubating cell suspensions of *K. stuttgartiensis* with ammonium, nitrite (2 mM each) and 100 μM NO scavenger PTIO (2-phenyl-4,4,5,5-tetramethylimidazoline-1-oxyl-3-oxide)¹². When PTIO was introduced at the start of the incubation or when it was added to active cells, anammox activity was inhibited (Fig. 2a). Further, the cells were incubated with ammonium and nitrite (2 mM each) in the presence of DAF2-DA (10 μM) that reacts with NO to form a fluorescent product^{13,14}. Sampled *K. stuttgartiensis* cells displayed the characteristic green fluorescence indicating NO production (Fig. 2b and Supplementary Fig. 1). In control experiments without nitrite or with added PTIO, there was no detectable fluorescent signal. It should be noted that both PTIO and DAF2-DA might have a wider reaction spectrum than NO and might

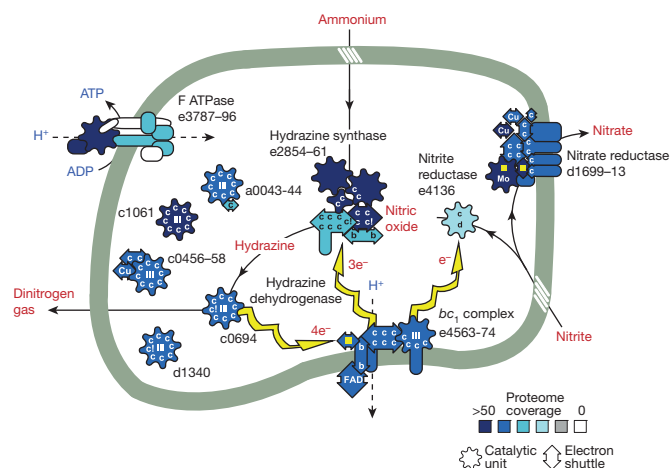


Figure 1 | Biochemical pathway and enzymatic machinery of *K. stuttgartiensis*. The anammoxosome, an intracytoplasmic compartment bounded by a membrane (grey line), is the locus of anammox catabolism. Identifiers of open reading frames and the degree to which the encoded respiratory protein complexes were detected in the proteome are indicated. Hydrazine synthase depicted in the centre of the figure is also loosely membrane associated. Yellow arrows, electron flow; yellow square, iron-sulphur clusters; b, haem b; c, haem c; cl, atypical haem c; d, haem d; Mo, molybdopterin. Cofactors and motifs were determined previously⁶.

¹Institute for Water and Wetland Research, Department of Microbiology, Radboud University Nijmegen, Heyendaalseweg 135, 6525AJ Nijmegen, The Netherlands. ²Nijmegen Centre for Mitochondrial Disorders, Nijmegen Proteomics Facility, Department of Laboratory Medicine, Radboud University Nijmegen Medical Centre, Geert Grooteplein 10, 6500HB Nijmegen, The Netherlands. ³Radboud University, Department of Molecular Biology, Nijmegen Centre for Molecular Life Sciences, Geert Grooteplein 28, 6525 GA Nijmegen, The Netherlands. ⁴Delft University of Technology, Department of Biotechnology, 2628 BC Delft, The Netherlands. ⁵Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, 28359 Bremen, Germany.

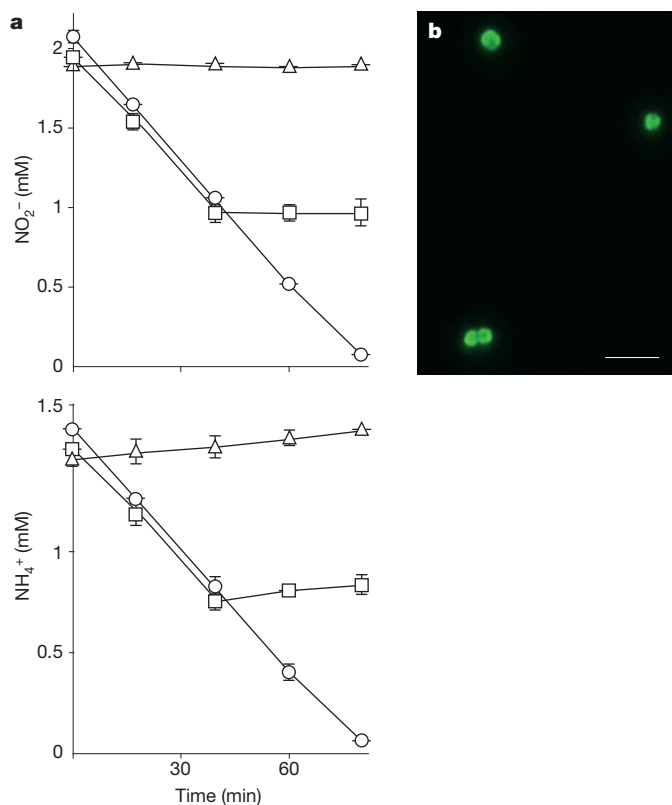
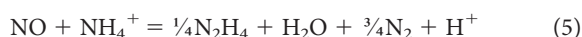


Figure 2 | Determination of nitric oxide (NO) as an intermediate. NO_2^- and NH_4^+ (2 mM each) conversion was inhibited by 100 μM PTIO (a). PTIO added at $t = 0$ (open triangle), PTIO added at 40 min (open square) and without PTIO (open circle), $n = 2$ (error bars, s.d.). (b) Epifluorescence image of (diaminofluorescein-2-diacetate) DAF2-DA derivative of NO formed during NH_4^+ and NO_2^- (2 mM each) conversion by anammox bacteria (scale bar, 10 μm).

possibly react with other nitrogen monoxides such as nitroxyl (HNO). However, unlike NO, HNO was not a suitable substrate for hydrazine synthase (see below).

Interestingly, when acetylene (15 μM) was added, the anammox reaction was inhibited. Acetylene inhibits aerobic ammonia monooxygenase, the ammonia-activating enzyme of aerobic ammonium oxidizers^{15–17}. Apparently, it also interfered with the ammonium-activating step of anammox cells (equation (2)). Importantly, acetylene inhibition resulted in an immediate accumulation of NO; hydroxylamine accumulation was not observed, consistent with the role of NO as the direct precursor for N_2H_4 .

The second step of the predicted anammox pathway would then be the reduction of NO and its simultaneous condensation with ammonium to produce N_2H_4 (equation (2)). Because the role of N_2H_4 in anammox catabolism was not established, we first demonstrated its *in vivo* turnover (Fig. 3a, b). To investigate whether N_2H_4 could be produced directly from NO, cell suspensions were incubated with NO (0.1 mM) and ammonium (2 mM). A transient accumulation of hydrazine (18 μM) was observed (Fig. 3c), albeit at a much lower concentration (200–500 μM) for incubations with hydroxylamine^{7,8}. This is consistent with equations (1)–(3) because the major part of the produced N_2H_4 would be oxidized to N_2 as expected from the overall reaction and NO could be supplied at much lower concentrations (equation (5)).



The anammox pathway is completed by the oxidation of N_2H_4 to N_2 (equation (3)). For a long time, N_2H_4 was known as an alternative

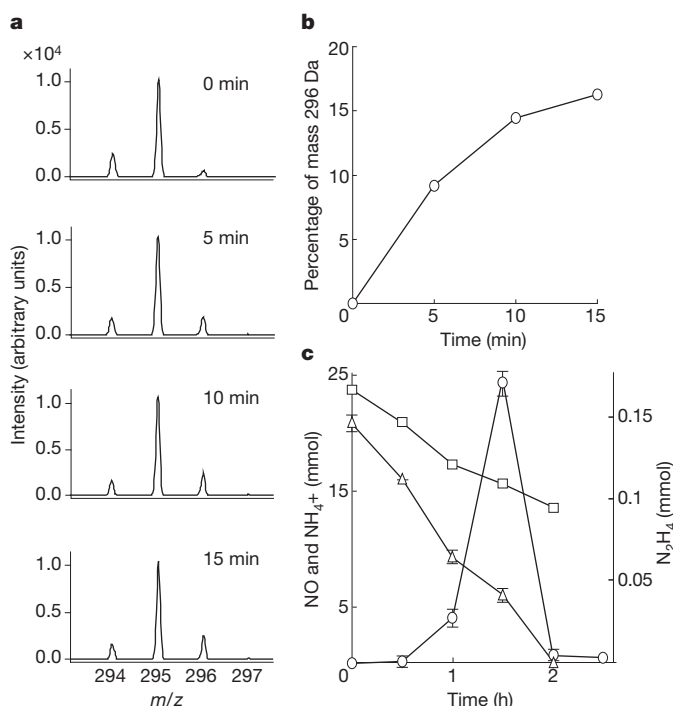


Figure 3 | Hydrazine turnover. *K. stuttgartiensis* cells were incubated with 2 mM $^{15}\text{NO}_2^-$ and $^{14}\text{NH}_4^+$ each in the presence of 2 mM $^{28}\text{N}_2\text{H}_4$. Under these conditions cells would only produce $^{29}\text{N}_2\text{H}_4$ and preferentially consume $^{28}\text{N}_2\text{H}_4$, leading to ^{15}N -label accumulation in the N_2H_4 pool. The 295 and 296 m/z masses correspond to derivatization products of $^{28}\text{N}_2\text{H}_4$ and $^{29}\text{N}_2\text{H}_4$ with *para*-dimethylaminobenzaldehyde³⁰ (a). The 294 m/z mass arises from the impurities of the matrix. Within 15 min, 16% of the N_2H_4 pool was labelled (b). Hydrazine (open circles) was produced by the cells incubated with 2 mM NH_4^+ (open triangles) and NO (0.1 mM) (open squares), $n = 2$ (error bars, s.d.) (c).

substrate for octahaem hydroxylamine oxidoreductases (HAOs), the enzymes that catalyse the conversion of hydroxylamine to nitrite in aerobic ammonium oxidizers^{18,19}. Strikingly, the *K. stuttgartiensis* genome encoded ten divergent paralogues of this enzyme, and six were detected at high levels in the transcriptome and proteome (mRNA up to 189-fold coverage, 27–58% p.p.d.; Supplementary Table 1). Six expressed paralogues belonged to the ‘type II’ hydrazine/hydroxylamine oxidoreductases (HZO/HAO)²⁰. Two related ‘type II’ HZO/HAO and one divergent octahaem cytochrome *c* were also detected at lower levels (4–15% p.p.d.) and one was not detected. By a two-step liquid chromatography procedure, we purified two highly expressed HZO/HAO-like proteins (kustc0694 and kustc1061). These enzymes appear to be closely related to two enzymes of unknown function isolated from an anammox enrichment culture KSU-1 (refs 21, 22). Both enzymes catalysed the four-electron oxidation of N_2H_4 to N_2 with cytochrome *c* as the artificial electron acceptor with different rates (2.5 and 0.4 $\mu\text{mol min}^{-1} \text{mg protein}^{-1}$, respectively). When they were incubated with $^{30}\text{N}_2\text{H}_4$ and cytochrome *c*, $^{30}\text{N}_2$ ($^{15}\text{N}^{15}\text{N}$) was produced stoichiometrically, in agreement with equation (3). Interestingly, Kustc1061 also oxidized hydroxylamine to NO (rather than nitrite) with a higher rate (6 $\mu\text{mol min}^{-1} \text{mg protein}^{-1}$). In contrast, kustc0694 did not catalyse this reaction and hydroxylamine and NO were powerful inhibitors of N_2H_4 oxidation, suggesting kustc0694 was the dedicated hydrazine dehydrogenase (HDH) in *K. stuttgartiensis*. Furthermore, the inhibition of kustc0694 explained the transient accumulation of N_2H_4 in the presence of hydroxylamine or NO.

Although no enzyme is known to convert NO and ammonium into N_2H_4 , two candidate gene clusters were previously identified potentially encoding an enzyme complex with this function (hydrazine synthase, HZS)⁶. One of these clusters (kuste2859–61) encoded the

most highly expressed proteins in the proteome (greater than 60% p.p.d., visible as three dominant spots on two-dimensional gels; Supplementary Table 1 and Supplementary Fig. 2a) and extremely abundant mRNAs in the transcriptome (greater than 50-fold coverage). The transcription of the other candidate cluster (kuste2474–83) was well below average (1.7-fold coverage) and expression was not detected by proteomics.

The kuste2859–61 proteins were purified from the cell-free extract of the *K. stuttgartiensis* as a complex that separated into three distinct bands on a denaturing polyacrylamide gel, corresponding to polypeptides encoded by three consecutive genes (kuste2859–2860–2861, Supplementary Fig. 2). Native polyacrylamide gel electrophoresis revealed that the complex was a multimer of approximately 240 kDa. Hydrazine synthesis activity of the complex was shown in a coupled assay with the kusc1061 HZO/HAO, using ^{15}N -ammonium (1 mM) and NO (0.9 mM) as substrates (Fig. 4). In the assay, kusc1061 would ‘pull’ the reaction by rapidly oxidizing the produced N_2H_4 to $^{29}\text{N}_2$ as the end product, while simultaneously ‘pushing’ the reaction by providing the electrons for N_2H_4 synthesis (equations (2) and 3). Kusc1061 alone did not catalyse the reaction, and N_2 production above background could not be measured in the absence of ammonium or NO. N_2 was not produced above background when hydroxylamine or nitroxyl (HNO) were provided as substrates with ammonium. The activity of N_2 formation in the coupled assay was $20 \text{ nmol h}^{-1} \text{ mg protein}^{-1}$, lower than the activity of whole cells with ammonium and nitrite (approximately $1800 \text{ nmol h}^{-1} \text{ mg protein}^{-1}$). The cell-free extracts were unable to form N_2 from ammonium and nitrite, but could from NO and ammonium under the same experimental conditions, at six-fold lower rate than the purified HZS ($3.4 \text{ nmol h}^{-1} \text{ mg protein}^{-1}$). The decrease in activity upon mere cell disruption was most probably due to the disruption of a tightly coupled multi-component system with hydrazine synthesis as the rate-limiting step.

Interestingly, the kuste2859–61 complex was capable of N_2 formation from ammonium and NO on its own (Fig. 4). The purified enzyme oxidized N_2H_4 to N_2 with a specific activity of $34 \text{ nmol min}^{-1} \text{ mg protein}^{-1}$, resulting in an overall disproportionation reaction (equation (5)). Considering that N_2H_4 is the energy source in anammox

metabolism, N_2 formation by HZS would be unproductive. Consequently, we may speculate that the anammox bacterium harbours backup systems that efficiently trap hydrazine and that keep (nitrogenous) inhibitory compounds, like NO and hydroxylamine, at low concentrations, which would partly explain the redundancy of HAO/HZO-like proteins in the organism. Our experiments showed that HZS and HDH were necessary and sufficient to make N_2 from the substrate ammonium and the intermediate NO.

Taken together, anammox catabolism and energy for growth must be conserved from three reactions (equations (1)–(3)). It is hypothesized that anammox bacteria synthesize ATP through a membrane-bound ATP synthase complex driven by proton-motive force (pmf) generated through catabolic reactions with the intermediary action of the quinol:cytochrome *c* oxidoreductase system (complex III, the bc_1 complex).

Intiguously, three gene clusters encoding bc_1 complexes and four encoding ATP synthases were present in the *K. stuttgartiensis* genome. Transcription and expression of one (kuste4569–74) of these gene clusters were detected at higher levels (26–33% p.p.d., 6- to 24-fold mRNA coverage) than the other two (0–19% p.p.d., 2- to 15-fold mRNA coverage). When *K. stuttgartiensis* cell suspensions were spiked with pentachlorophenol (10 μM), a structural analogue of quinol and a known inhibitor of the bc_1 complex, anammox activity was completely inhibited, indicating that the bc_1 complex was involved in energy conservation and its role in electron transport from N_2H_4 oxidation to nitrite reduction and hydrazine synthesis was not backed up by any other system. The expression of the four gene clusters encoding ATP synthase was even more skewed. Peptide coverage for kuste3789–96 was 14–58% p.p.d. compared with less than 1% for the other three ATP synthases, and mRNA coverage differed by a factor of six. The gene product encoding the catalytic β -subunit of the highest expressed ATP synthase (kuste3787–96) was recently shown to be associated with the membranes of the intracellular cell compartment, the anammoxosome, suggesting it to be the site where the proton-motive machinery resides²³.

In the present study we experimentally identified NO and N_2H_4 as the intermediates of anaerobic ammonium oxidation. The highly expressed protein encoded by the gene cluster kuste2859–61 was purified and N–N bond formation from NO and ammonium was demonstrated. Hydrazine synthase and the NO reductase of denitrifiers are the two enzymes capable of bonding two N atoms together. In contrast to NO reductase, hydrazine synthase combines two different nitrogenous molecules. It is intriguing that all the N_2 in our atmosphere is formed by the oxidizing power of NO, in line with the hypothesis that NO may have been the first deep redox sink on Earth²⁴.

METHODS SUMMARY

Activity Measurements. Physiological experiments were performed at 33 °C, pH 7.5 with *K. stuttgartiensis* cells^{9,10}. To determine the role of NO and hydroxylamine in the anammox metabolism, cells were incubated with (1) NaNO_2 , NH_4Cl (2 mM each) and spiked with acetylene (15 μM); (2) NaNO_2 , NH_4Cl (2 mM each) and DAF-2DA (10 μM) or PTIO (100 μM); (3) NO (0.1 mM) and 2 mM NH_4Cl . Hydroxylamine, NH_4^+ , NO_2^- and N_2H_4 were determined as previously described^{25,26}. NO was measured online as previously described²⁷. To determine N_2H_4 turnover, cells were incubated with $\text{Na}^{15}\text{NO}_2$, NH_4Cl (or vice versa) and N_2H_4 (2 mM each). Isotopic composition of hydrazine was determined with matrix-assisted laser desorption/ionization–time of flight mass spectrometry (MALDI–TOF MS) after *para*-dimethylaminobenzaldehyde derivatization, developed after Watt and Chrisp²⁵. All labelled compounds were 99% pure (Cambridge Isotope Laboratories).

Proteins were purified from cell-free extracts with anion exchange and hydroxyapatite liquid chromatography. Activity measurements were performed at 37 °C, pH 7 in an anaerobic chamber. Kuste2859–61 (1.6 mg) and kusc1061 (4.7 μg) were incubated with NO (0.9 mM) and $^{15}\text{NH}_4\text{Cl}$ (1 mM), and $^{29}\text{N}_2$ production was monitored by gas chromatography (Agilent 6890 with a PorapakQ column, 80 °C) combined with a mass spectrometer (Agilent 5975c quadrupole inert MS). For rate calculations, kusc0694 (1.3 μg) or kusc1061 (4.7 μg) were

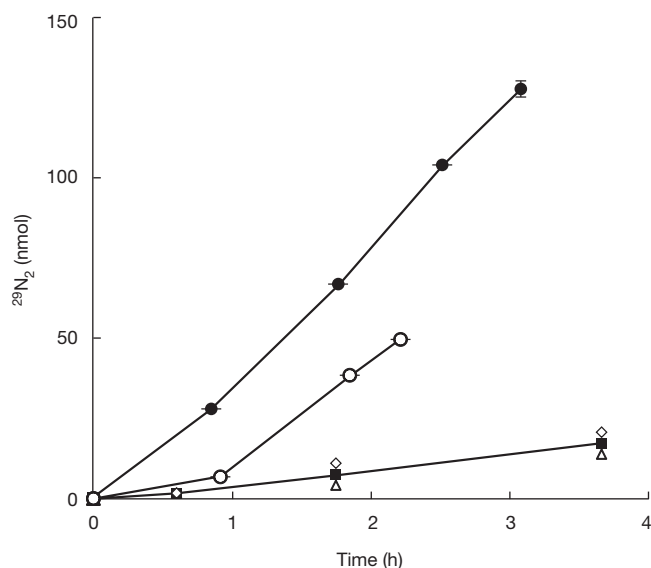


Figure 4 $^{29}\text{N}_2$ production by hydrazine synthase complex and kusc1061 from $^{15}\text{NH}_4^+$ and NO. $^{29}\text{N}_2$ was produced with the highest rate when hydrazine synthase complex (1.6 mg) and kusc1061 (4.7 μg) was incubated with $^{15}\text{NH}_4^+$ (1 mM), ^{14}NO (0.9 mM) and cytochrome *c* (50 μM) (filled circles). In the control experiments, hydrazine synthase complex and cytochrome *c* (open circles), kusc1061 and cytochrome *c* (open diamonds), cytochrome *c* (filled squares) and only buffer (open triangles) were incubated under the same experimental conditions; $n = 3$ (error bars, s.d.).

incubated with N_2H_4 or hydroxylamine and cytochrome *c* (50 μM each), and $^{30}\text{N}_2$ or ^{31}NO production was measured.

Molecular methods. Total RNA was extracted, reverse transcribed, sequenced with Illumina and mapped to the genome sequence of *K. stuttgartiensis*⁶. From the aligned reads, per-position coverage was calculated for each contig and used to calculate the coverage for each orf, intergenic region and predicted RNA element.

Cell free extracts were separated by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) or two-dimensional gel electrophoresis, digested with trypsin and analysed with liquid chromatography–mass spectrometry (LC–MS/MS)^{28,29}. Mass spectrometry data was searched against a database of predicted *K. stuttgartiensis* peptide sequences.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 May; accepted 12 August 2011.

Published online 2 October 2011.

- Arrigo, K. R. Marine microorganisms and global nutrient cycles. *Nature* **437**, 349–355 (2005).
- Brandes, J. A., Devol, A. H. & Deutsch, C. New developments in the marine nitrogen cycle. *Chem. Rev.* **107**, 577–589 (2007).
- Payne, W. J. Reduction of nitrogenous oxides by microorganisms. *Bacteriol. Rev.* **37**, 409–452 (1973).
- Strous, M. *et al.* Missing lithotroph identified as new planctomycete. *Nature* **400**, 446–449 (1999).
- Zumft, W. G. Cell biology and molecular basis of denitrification. *Microbiol. Mol. Biol. Rev.* **61**, 533–616 (1997).
- Strous, M. *et al.* Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* **440**, 790–794 (2006).
- Van de Graaf, A. A., deBruijn, P., Robertson, L. A., Jetten, M. S. M. & Kuenen, J. G. Metabolic pathway of anaerobic ammonium oxidation on the basis of N-15 studies in a fluidized bed reactor. *Microbiology* **143**, 2415–2421 (1997).
- Kartal, B. *et al.* Candidatus ‘Brocadia fulgida’: an autofluorescent anaerobic ammonium oxidizing bacterium. *FEMS Microbiol. Ecol.* **63**, 46–55 (2008).
- Kartal, B., Geerts, W. & Jetten, M. S. M. Cultivation, detection and ecophysiology of anaerobic ammonium-oxidizing bacteria. *Methods Enzymol.* **486**, 89–108 (2011).
- Van der Star, W. R. L. *et al.* The membrane bioreactor: a novel tool to grow anammox bacteria as free cells. *Biotechnol. Bioeng.* **101**, 286–294 (2008).
- Keller, M. & Hettich, R. Environmental proteomics: a paradigm shift in characterizing microbial activities at the molecular level. *Microbiol. Mol. Biol. Rev.* **73**, 62–70 (2009).
- Akaike, T. & Maeda, H. Quantitation of nitric oxide using 2-phenyl-4,4,5,5-tetramethylimidazole-1-oxyl 3-oxide (PTIO). *Methods Enzymol.* **268**, 211–221 (1996).
- Guo, F. Q., Okamoto, M. & Crawford, N. M. Identification of a plant nitric oxide synthase gene involved in hormonal signaling. *Science* **302**, 100–103 (2003).
- Nagano, T. Practical methods for detection of nitric oxide. *Luminescence* **14**, 283–290 (1999).
- Gilch, S., Vogel, M., Lorenz, M. W., Meyer, O. & Schmidt, I. Interaction of the mechanism-based inactivator acetylene with ammonia monooxygenase of *Nitrosomonas europaea*. *Microbiology* **155**, 279–284 (2009).
- Hyman, M. R. & Wood, P. M. Suicidal inactivation and labeling of ammonia monooxygenase by acetylene. *Biochem. J.* **227**, 719–725 (1985).
- McTavish, H., Fuchs, J. A. & Hooper, A. B. Sequence of the gene coding for ammonia monooxygenase in *Nitrosomonas europaea*. *J. Bacteriol.* **175**, 2436–2444 (1993).
- Hooper, A. B. & Nason, A. Characterization of hydroxylamine-cytochrome *c* reductase from chemoautotrophs *Nitrosomonas europaea* and *Nitrosocystis oceanus*. *J. Biol. Chem.* **240**, 4044–4057 (1965).
- Hooper, A. B., Vannelli, T., Bergmann, D. J. & Arciero, D. M. Enzymology of the oxidation of ammonia to nitrite by bacteria. *Anton. Leeuw. Int. J. G. Microbiology* **71**, 59–67 (1997).
- Klotz, M. G. *et al.* Evolution of an octahaem cytochrome *c* protein family that is key to aerobic and anaerobic ammonia oxidation by bacteria. *Environ. Microbiol.* **10**, 3150–3163 (2008).
- Shimamura, M. *et al.* Isolation of a multiheme protein with features of a hydrazine-oxidizing enzyme from an anaerobic ammonium-oxidizing enrichment culture. *Appl. Environ. Microbiol.* **73**, 1065–1072 (2007).
- Shimamura, M. *et al.* Another multiheme protein, hydroxylamine oxidoreductase, abundantly produced in an anammox bacterium besides the hydrazine-oxidizing enzyme. *J. Biosci. Bioeng.* **105**, 243–248 (2008).
- Van Niftrik, L. *et al.* Intracellular localization of membrane-bound ATPases in the compartmentalized anammox bacterium ‘Candidatus Kuenenia stuttgartiensis’. *Mol. Microbiol.* **77**, 701–715 (2010).
- Ducluzeau, A. L. *et al.* Was nitric oxide the first deep electron sink? *Trends Biochem. Sci.* **34**, 9–15 (2009).
- Watt, G. W. & Chrisp, J. D. A spectrophotometric method for the determination of hydrazine. *Anal. Chem.* **24**, 2006–2008 (1952).
- Strous, M., Heijnen, J. J., Kuenen, J. G. & Jetten, M. S. M. The sequencing batch reactor as a powerful tool for the study of slowly growing anaerobic ammonium-oxidizing microorganisms. *Appl. Microbiol. Biotechnol.* **50**, 589–596 (1998).
- Kartal, B. *et al.* Effect of nitric oxide on anammox bacteria. *Appl. Environ. Microbiol.* **76**, 6304–6306 (2010).
- Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
- Wilm, M. *et al.* Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**, 466–469 (1996).
- Audrieth, L. F. & Ackerson Ogg, B. *The Chemistry of Hydrazine* (Wiley, 1951).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements B.K. was supported by a grant (05987) from the Dutch Foundation for Applied Research. W.J.M. was supported by a grant (142161201) from the Darwin Center for Biogeosciences. N.M.d.A. and I.C. were supported by a grant (81802015) from the Netherlands Organization for Scientific Research. M.S. was supported by a VIDI grant from the Netherlands Organization for Scientific Research and a European Research Council grant MASEM (242635). The anammox research of M.S.M.J. is supported by an advanced grant (232987) from the ERC. The authors acknowledge R. Klefoth for the initial tests for protein purification procedures.

Author Contributions Physiological experiments were conceived, designed and performed by B.K., *Kuenenia stuttgartiensis* was grown by B.K. and W.G., two-dimensional gel electrophoresis was performed by N.M.A. and I.C., one-dimensional gel electrophoresis was performed by W.J.M. and B.K., MALDI–TOF analysis was performed by B.K., W.J.M. and H.J.M.O.d.C., nanoLC–MS/MS by J.G., RNA extraction and reverse transcription by H.R.H., Illumina sequencing by E.M.J.-M., K.-J.F. and H.S., and protein purification and activity tests were designed by W.J.M., B.K. and J.T.K. and performed by W.J.M. Proteomic and transcriptomic data processing was performed by J.G., M.S., K.-J.F., B.K., M.S.M.J. and H.J.M.O.d.C. The manuscript was written by B.K. with input from J.T.K., M.S. and M.S.M.J.

Author Information The metatranscriptome and peptidome sequences are deposited in Gene Expression Omnibus under accession numbers GSE15408 and PSE111, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.K. (kartal@science.ru.nl).

METHODS

Source of the biomass. *K. stuttgartiensis* cells were collected from a 10-l laboratory scale anammox membrane bioreactor^{9,10} and were concentrated by centrifugation. The cells were re-suspended to a protein concentration higher than 1 mg ml⁻¹. Part of the cell suspension was diluted 100 times, chemically fixed, and hybridizations with fluorescently labelled oligonucleotide probes were performed as described previously^{31,32}.

Sample preparation. The cell suspensions were transferred to 8-ml serum bottles. The vials were made anoxic by alternately applying under-pressure and He or Ar seven times and were transferred to an anaerobic chamber with a 95%/5% Ar/H₂ atmosphere. O₂ in the Ar in the anaerobic chamber was removed by passing Ar over a Pd catalyst (0.2 p.p.m. residual O₂). In the anaerobic chamber, cell suspensions were diluted five times with anaerobic mineral medium³³ (pH 7.5) to a final volume of 40 or 8 ml and transferred to glass vials unless stated otherwise. All preparations (for example, addition of substrates and/or inhibitors) for different incubations were handled in the anaerobic chamber. All experiments were performed at least in duplicate. All non-labelled salts were purchased as molecular grade (more than 99.95% pure, Merck) unless stated otherwise. All labelled compounds were 99% pure and purchased as sodium or chloride salts (Cambridge Isotope Laboratories). All gaseous compounds were of the highest purity available.

Analytical methods. NO₂⁻ and NH₄⁺ were determined as described previously²⁷. N₂H₄ was determined colourimetrically at 420 nm after reaction of 100-μl sample with 900 μl 2% (w/v) *para*-dimethylaminobenzaldehyde (PDB), 3.7% (v/v) HCl in ethanol³⁰. NH₂OH (detection limit 5 μM) was determined as previously described²⁷.

Effect of PTIO. To determine the effect of PTIO, an NO scavenger¹², on anammox bacteria, three incubations were performed in parallel. NO₂⁻ and NH₄⁺ (2 mM) were added to all incubations. To the first incubation, PTIO (100 μM) was added at 0 min, to the second it was added at 40 min, and no PTIO was added to the third incubation. Liquid samples were taken every 15 min and analysed for NH₄⁺, NO₂⁻ and NH₂OH as previously described²⁷.

Bioimaging of nitric oxide. To detect NO turnover, *K. stuttgartiensis* cell suspensions were incubated with 2 mM NO₂⁻ and NH₄⁺ in amber vials. In parallel, nitrite-depleted cell suspensions were incubated in the presence of 2 mM NH₄⁺. After a 5-min pre-incubation, diamino fluorescein-2-diacetate (DAF2-DA, Calbiochem) was added to a final concentration of 10 μM. The vials were incubated in the dark for 30 min at 33 °C and were shaken continuously at 300 r.p.m. As a negative control, cells were incubated with PTIO and DAF2-DA. Cells were then harvested by centrifugation, washed three times in mineral medium³³ to remove the excess chromophore and were re-suspended in mineral medium. A liquid sample (5 μl) of the suspension was pipetted on a microscope slide and dried in the dark. The preparations were examined with a Zeiss Axioplan2 epifluorescence microscope.

Batch experiments. To determine the activity of *K. stuttgartiensis* with NO and NH₄⁺, cell suspensions were incubated with NO (0.1 mM) and 2 mM NH₄⁺ in 100-ml glass vials with 10% NO (in He) in the headspace. Gas samples were analysed in a chemoluminescence NO_x analyser (CLD 700EL, EcoPhysics, detection limit 0.1 p.p.m. NO). Liquid samples were taken every 30 min and analysed for NH₄⁺ and N₂H₄ as previously described^{27,30}.

To determine the effect of acetylene on anammox bacteria, *K. stuttgartiensis* suspensions were transferred to 40-ml glass vials. NO₂⁻ and NH₄⁺ were added to the incubations to a final concentration of 2 mM. The vials were incubated at 33 °C and were mixed with a magnetic stirrer at 500 r.p.m. and continuously flushed with Ar/CO₂ (95%/5%) with a flow of 10 ml min⁻¹. The effluent gas from the vials was connected to a chemoluminescence NO_x analyser (CLD 700EL, EcoPhysics, detection limit 0.1 p.p.m. NO) for online NO measurement. At 15 min, 100 μl acetylene (15 μM) was added to the vials. As negative controls, 100 μl air and 100 μl nitrogen were added to separate incubations. Liquid samples were taken from the incubations every 10–15 s and chilled to 0 °C immediately. The supernatant of each sample was transferred to an Eppendorf cup and kept at 4 °C until they were analysed for NH₄⁺, N₂H₄ and NH₂OH.

Detection of hydrazine turnover in anammox cells. To detect N₂H₄ turnover, ¹⁵NO₂⁻, NH₄⁺ (or vice versa) and N₂H₄ (2 mM each) were added to the *K. stuttgartiensis* cell suspensions. The vials were incubated in the dark for 15 min at 30 °C, 300 r.p.m. Liquid samples were taken every 5 min, and isotopic composition of N₂H₄ was determined by MALDI-TOF MS after reaction with PDB. For MALDI-TOF analysis, 10 μl of PDB-reacted samples were mixed with an equal volume of sample buffer containing 20 mg ml⁻¹ α-cyano-4-hydroxycinnamic acid in 0.05% (v/v) trifluoroacetic acid (TFA), 50% (v/v) acetonitrile. The mixtures (0.3 μl) were spotted on a S26/100 M-probe (Bruker 15165), which was inserted into a multiprobe adaptor. MALDI-TOF MS measurements were performed in the mass range of 100–800 Da on a Bruker III mass spectrometer, using the reflectron mode.

Cytochrome *bc*₁ complex. To determine the role of cytochrome *bc*₁ complex in anammox catabolism, *K. stuttgartiensis* suspensions were incubated with pentachlorophenol, a specific inhibitor of the *bc*₁ complex. NO₂⁻ and NH₄⁺ were added to the incubations to a final concentration of 2 mM. Pentachlorophenol was added to a final concentration of 10 μM. NO₂⁻ and NH₄⁺ were determined as described previously²⁷.

Preparation of cell free extract. *K. stuttgartiensis* cells (2 l, OD₆₀₀ 1.2) were harvested from the membrane bioreactor. After centrifugation (4,000g, 4 °C), the pellet was re-suspended in one volume 20 mM potassium phosphate buffer, pH 8. Cell suspensions were passed three times through a French pressure cell operated at 138 MPa. The lysate was incubated with 1% (w/v) sodium deoxycholate at 4 °C for 1 h to solubilize membrane associated proteins. After centrifugation for 15 min at 1,700g at 4 °C, the cell-free fraction was obtained as clarified supernatant.

Protein electrophoresis and MALDI-TOF analysis. Samples were denatured by incubation with 60 mM Tris-HCl buffer (pH 8) containing 5% β-mercaptoethanol, 2% SDS (sodium dodecyl sulphate) and 25% glycerol for 5 min at 100 °C. SDS-PAGE was performed in 10% or 6% slab gels in 375 mM Tris-HCl glycine buffer, pH 8.8 according to Laemmli³⁴. Native PAGE (6%) was performed according to the same procedure with the following modifications: the protein preparations were not boiled before electrophoresis, SDS and β-mercaptoethanol were omitted from the gels, and Tris-HCl glycine (375 mM, pH 8.3) was used as the running buffer. Gels were stained with colloidal Coomassie blue as described elsewhere³⁵. To identify the protein bands resolved in SDS-PAGE, gel spots (~3 mm²) were picked, digested with trypsin and analysed with MALDI-TOF mass spectrometry as described elsewhere³⁶.

Purification of kuste2859–2860–2861, kusc0694 and kusc1061. Cell-free extract was centrifuged at 140,000g, 10 °C (Discovery 10, Sorvall, equipped with a T-1270 rotor) to remove the membranes. The supernatant was loaded on a 30 ml Q Sepharose XL (GE Healthcare) column equilibrated with 20 mM Tris-HCl, pH 8. Kuste2859–2860–2861 and kusc1061 were eluted isocratically with 200 mM NaCl in 20 mM Tris-HCl, pH 8 (2 ml min⁻¹). Kusc0694 was eluted isocratically with 400 mM NaCl in 20 mM Tris-HCl, pH 8 (2 ml min⁻¹). Eluted fractions were subsequently loaded onto a 10 ml Hydroxapate (Bio-Rad) column equilibrated with 20 mM potassium phosphate buffer, pH 7 and eluted with a gradient of the same buffer (20–500 mM, 2 ml min⁻¹). Kusc1061 and kuste2859–2860–2861 were collected in fractions eluted at 100 mM and 200 mM phosphate, respectively. The pooled fractions were desalted and concentrated using Vivaspins tubes (100 kDa cut-off, Sartorius Stedim Biotech) to concentrations of at least 0.86 mg ml⁻¹ (kuste2859–2860–2861) and 2.3 mg ml⁻¹ (kusc1061) in 20 mM phosphate buffer, pH 7.

Detection of hydrazine and hydroxylamine oxidation by kusc1061 and kusc0694. To 2 ml (final volume) of phosphate buffer (20 mM, pH 7), 4.7 μg of Kusc1061 or 1.3 μg of Kusc0694 and cytochrome *c* (50 μM final concentration, bovine heart, Sigma-Aldrich) were added to a 3-ml exetainer (Labco). To start the reaction to determine the electron stoichiometry, 10 μM, and for routine rate assays 50 μM, ¹⁵N-labelled ³⁰N₂H₄ was added from an anoxic stock. To determine the capacity for NH₂OH oxidation, proteins were incubated in separate vials with 50 μM NH₂OH and cytochrome *c* (each). Exetainers were incubated at 37 °C in the anaerobic chamber. ³⁰N₂ and ¹⁵NO production was monitored by gas chromatography (Agilent 6890 equipped with a Porapak Q column at 80 °C) combined with a mass spectrometer (Agilent 5975c quadrupole inert MS).

Combined assay of kuste2859–2860–2861 and kusc1061. Cytochrome *c* (50 μM final concentration, bovine heart, Sigma-Aldrich), Kusc1061 (4.7 μg), 1 mM ¹⁵NH₄⁺ and 5 μM N₂H₄ were added to 1.6 mg of kuste2859–2860–2861 in 1 ml phosphate buffer (20 mM, pH 7) in a 3-ml exetainer (Labco). The reaction was started by adding phosphate buffer (20 mM, pH 7) with NO (0.9 mM) to a final volume of 2 ml. Before incubation at 37 °C in the anaerobic chamber, 1 ml of 50% NO (in He) was added to the headspace. Control experiments were performed with ammonium (1 mM) with NH₂OH (1 mM) and HNO supplied as Angeli's salt (41 mM) in separate incubations. ²⁹N₂ production was monitored by gas chromatography (Agilent 6890 equipped with a Porapak Q column at 80 °C) combined with a mass spectrometer (Agilent 5975c quadrupole inert MS).

LC-MS/MS analysis and data processing. After PAGE, gels were stained with colloidal Coomassie blue as described elsewhere³⁵. The gel lane was cut into four slices and each slice was destained with three cycles of washing with 50 mM ammoniumbicarbonate and 50% acetonitrile. Protein reduction, alkylation and digestion with trypsin were performed as previously described³⁰. After digestion, samples were de-salted and purified according to Rappsilber *et al.*²⁹. Sample analysis by LC-MS/MS was performed using an Agilent nanoflow 1100 liquid chromatograph coupled online through a nano-electrospray ion source (Thermo Fisher Scientific) to a 7T linear ion trap Fourier transform ion cyclotron resonance mass spectrometer (LTQ FT, Thermo Fisher Scientific). The chromatographic column consisted of a 15-cm fused silica emitter (New Objective, PicoTip

Emitter, Tip: $8 \pm 1 \mu\text{m}$, internal diameter $100 \mu\text{m}$) packed with $3\text{-}\mu\text{m}$ C18 beads (Reprosil-Pur C18 AQ, Dr Maisch GmbH)³⁷. After loading the peptides onto the column in buffer A (0.5% HAc), bound peptides were gradually eluted using a 67-min gradient of buffer B (80% ACN, 0.5% HAc). First, the concentration of acetonitrile was increased from 2.4 to 8% in 5 min, followed by an increase from 8 to 24% acetonitrile in 55 min, and finally an increase from 24 to 40% acetonitrile in 7 min. The mass spectrometer was operated in positive ion mode and was programmed to analyse the top four most abundant ions from each precursor scan using dynamic exclusion. Survey mass spectra (350–2000 m/z) were recorded in the ion cyclotron resonance cell at a resolution of $R = 5\text{E}5$. Data-dependent collision-induced fragmentation of the precursor ions was performed in the linear ion trap (normalized collision energy 27%, activation $q = 0.250$, activation time 30 ms).

Mass spectrometric data files were searched against the *K. stuttgartensis* database (known contaminants like human keratins and trypsin were added to the database) using the database search program Mascot (Matrix Science, version 2.2). To obtain factors for the recalibration of precursor masses, initial searches were performed with a precursor ion tolerance of 50 p.p.m. Fragment ions were searched with 0.8-Da tolerance and searches allowed for one missed cleavage, carbamidomethylation (C) as fixed modification, and deamidation (NQ) and oxidation (M) as variable modifications. The results from these searches were used to calculate the m/z -dependent deviation, which was used to recalibrate all precursor m/z values. After recalibration of the precursor masses, definitive Mascot searches were performed using the same settings as stated above, but with a precursor ion tolerance of 20 p.p.m. Additionally, reverse database searches were performed with the same settings. Protein identifications were validated and clustered using the PROVALT algorithm to achieve a false-discovery rate of less than 1% (ref. 38).

Two-dimensional gel electrophoresis. Before protein separation by isoelectric focusing, 1 mg of the protein suspension was incubated with 1% (v/v) Immobilized pH-gradient (IPG) buffer of the appropriate range, 5 mM tributyl phosphine and 0.01% (w/v) bromophenol blue for 15 min at room temperature and centrifuged at 10,000g for 15 min at 10°C . Isoelectric focusing was performed with the IPGphor system using commercial 24-cm-long IPG strips with linear immobilized pH gradients of various ranges. The conditions for rehydration of the IPG strips, sample entry and isoelectric focusing were as follows: the temperature was set constant at 18°C and 50- μA per strip were applied.

Focused IPG strips were equilibrated before SDS-PAGE two times for 15 min in 375 mM Tris-HCl pH 8.5, 2% (w/v) SDS, 20% (w/v) glycerol, 6 M urea, 10 mM DTT, 50 mM acrylamide and 0.1% (w/v) bromophenol blue. Gels were run for 45 min with constant cooling to 18°C at 20 V, 40 W and subsequently at 40 V, 40 W until the bromophenol blue marker reached the end of the gel. Gels were fixed in 30% (v/v) ethanol and 10% (v/v) glacial acetic acid and were stained with colloidal Coomassie blue³⁹ or silver stain⁴⁰. Picked gel spots were digested and analysed with MALDI-TOF MS as described elsewhere³⁶.

Blue native PAGE. Blue native PAGE of the protein complexes was performed as described elsewhere⁴¹. For protein identification in two-dimensional gels, $16\text{ cm} \times 20\text{ cm}$ gels were self-casted according to Calvaruso *et al.* with the following

exception: 4–10% linear polyacrylamide gradient was used⁴¹. Sample additive (1.5 μl) (0.75 M 6-aminocaproic acid, 5% Serva Blue G) was added to 40 μg protein sample before loading the gel.

Electrophoresis was performed at 50 V until the migration front entered the resolving gel and then at 100 V until the migration front reached the end of the gel. Cathode and anode buffer for blue native PAGE were 50 mM Bis-Tris, pH 7.0, and 50 mM Tricine, 15 mM Bis-Tris, pH 7.0, respectively. Preparation of the first-dimension gel strip and assembly and casting of the second-dimension gel were performed as described elsewhere⁴¹ with the exception that the second-dimension cassette had the same thickness as the first dimension. No Coomassie blue was added to the cathode buffer.

Transcriptomics. RNA was extracted using the Ribopure Bacteria Kit (Ambion) according to the manufacturer's instructions. First-strand cDNA was synthesized with random primers using the RevertAid H Minus First Strand cDNA Synthesis Kit, and the second strand was synthesized using DNA polymerase and manufacturer's instructions (Fermentas).

The quality scores of the obtained Solexa reads (3.5 million) were converted to PHRED format and mapped with Maq (<http://maq.sourceforge.net>) to the five contigs that constitute the *K. stuttgartensis* genome (accession numbers CT030148, CT573071–4). From the aligned reads, the per-position coverage was calculated for each contig and used to calculate the coverage for each orf, intergenic region and predicted RNA element.

- Schmid, M. *et al.* Molecular evidence for genus level diversity of bacteria capable of catalyzing anaerobic ammonium oxidation. *Syst. Appl. Microbiol.* **23**, 93–106 (2000).
- Schmid, M. C. *et al.* Biomarkers for in situ detection of anaerobic ammonium-oxidizing (anammox) bacteria. *Appl. Environ. Microbiol.* **71**, 1677–1684 (2005).
- Van de Graaf, A. A., de Bruijn, P., Robertson, L. A., Jetten, M. S. M. & Kuenen, J. G. Autotrophic growth of anaerobic ammonium-oxidizing micro-organisms in a fluidized bed reactor. *Microbiology* **142**, 2187–2196 (1996).
- Laemmli, U. K. Cleavage of structural proteins during assembly of head of bacteriophage-T4. *Nature* **227**, 680–685 (1970).
- Candiano, G. *et al.* Blue silver: A very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* **25**, 1327–1333 (2004).
- Farhoud, M. H. *et al.* Protein complexes in the archaeon *Methanothermobacter thermotrophicus* analyzed by blue native/SDS-PAGE and mass spectrometry. *Mol. Cell. Proteomics* **4**, 1653–1663 (2005).
- Ishihama, Y., Rappsilber, J., Andersen, J. S. & Mann, M. Microcolumns with self-assembled particle frits for proteomics. *J. Chromatogr. A* **979**, 233–239 (2002).
- Weatherly, D. B. *et al.* A heuristic method for assigning a false-discovery rate for protein identifications from mascot database search results. *Mol. Cell. Proteomics* **4**, 762–772 (2005).
- Neuhoff, V., Arold, N., Taube, D. & Ehrhardt, W. Improved staining of proteins in polyacrylamide gels including isoelectric-focusing gels with clear background at nanogram sensitivity using coomassie brilliant blue G-250 and R-250. *Electrophoresis* **9**, 255–262 (1988).
- Mortz, E., Krogh, T. N., Vorum, H. & Gorg, A. Improved silver staining protocols for high sensitivity protein identification using matrix-assisted laser desorption/ionization-time of flight analysis. *Proteomics* **1**, 1359–1363 (2001).
- Calvaruso, M. A., Smeitink, J. & Nijtmans, L. Electrophoresis techniques to investigate defects in oxidative phosphorylation. *Methods* **46**, 281–287 (2008).

Derivation of haploid embryonic stem cells from mouse embryos

Martin Leeb¹ & Anton Wutz¹

Most animals are diploid, but haploid-only and male-haploid (such as honeybee and ant) species have been described¹. The diploid genomes of complex organisms limit genetic approaches in biomedical model species such as mice. To overcome this problem, experimental induction of haploidy has been used in fish^{2,3}. Haploid development in zebrafish has been applied for genetic screening². Recently, haploid pluripotent cell lines from medaka fish (*Oryzias latipes*) have also been established³. In contrast, haploidy seems less compatible with development in mammals^{4,5}. Although haploid cells have been observed in egg cylinder stage parthenogenetic mouse embryos⁶, most cells in surviving embryos become diploid. Here we describe haploid mouse embryonic stem cells and show their application in forward genetic screening.

Previous attempts to establish pluripotent stem-cell lines from haploid mouse embryos have resulted in the isolation of parthenogenetic embryonic stem (ES) cells with a diploid karyotype⁴. These studies reported the development of apparently normal haploid mouse blastocysts with a defined inner cell mass (ICM)^{4,5}. To investigate the haploid ICM, we cultured haploid mouse blastocysts in chemically defined medium with inhibitors of mitogen activated protein kinase kinase and glycogen synthase kinase 3. This 2i medium⁷ has previously been used for isolating ES cells from mouse strains that have been refractory to this process⁸ and from rats⁹, and may help to maintain certain characteristics of early mouse epiblast cells^{10,11}.

We generated haploid mouse embryos by activation of unfertilized oocytes isolated from superovulated B6CBAF1 hybrid female mice using strontium chloride. After culture in M16 medium, 30 blastocysts (22%) were obtained from 132 activated oocytes and used for ES cell derivation. After removal of the zona and trophectoderm, ICMs were cultured in gelatinized 96-well dishes in 2i medium in the presence of leukaemia inhibitory factor (LIF). Twenty-seven ES cell lines were obtained (93%). Individual ES cell lines were expanded and their DNA content was analysed by flow analysis using diploid ES cells as controls (Fig. 1a, b). In six ES cell lines, at least 10% of the cells had a haploid DNA content and the proportion of haploid cells could reach a conservatively estimated 60% (Fig. 1b). Further enrichment was achieved by flow sorting of cells with a haploid DNA content after staining with Hoechst 33342 (Fig. 1c). This allowed expansion of haploid ES cell lines for over 35 passages.

We further tested the requirements for deriving haploid mouse ES cells (Table 1). These experiments showed that removal of the trophectoderm by immunosurgery was not essential. Haploid ES cells could also be established using DMEM medium supplemented with knockout serum replacement and LIF, showing that derivation without kinase inhibitors is possible (Table 1 and Supplementary Fig. 1). We further succeeded in isolating haploid ES cells from the 129Sv inbred mouse strain and two genetically modified mouse lines. In the latter, several alleles had been bred to homozygosity and maintained on a mixed genetic background for several generations (Table 1 and Supplementary Fig. 2). In summary, we derived 25 haploid ES cell lines in 7 independent experiments. Haploid ES cell cultures could also

be maintained on feeders in serum containing DMEM supplemented with LIF.

Haploid ES cells exhibited a typical mouse ES cell colony morphology (Fig. 1d). Chromosome spreads showed 20 chromosomes corresponding to the haploid mouse chromosome set (Fig. 1e, f). For further characterizing the genetic integrity we performed comparative genomic hybridization (CGH) of four haploid ES cell lines and control DNA from the CBA strain and the mixed transgenic mouse line from which HTG-1 and HTG-2 ES cells were derived (Fig. 1g, h and Supplementary Figs 3 and 4). Copy number variations (CNVs) that were detected in the genome of haploid ES cell lines were also present in the strains of origin (Supplementary Table 1). Although some CNVs seemed to be specific to haploid ES cells, inspection of the actual signals (Supplementary Fig. 4) indicated that these CNVs were also present in the CBA or HTG control DNAs but not detected with the threshold applied. CNVs between the C57BL/6 and CBA strain of mice were consistent with a previously reported analysis¹². Taken together, these data show that haploid ES cells maintained an intact haploid genome without amplifications or losses.

At the molecular level, haploid ES cells expressed pluripotency markers including *Oct4* (also called *Pou5f1*), *Rex1* (also called *Zfp42*), *Klf4*, *Sox2* and *Nanog* (Fig. 2a, b). Genome-wide expression analysis showed a high correlation (Pearson correlation coefficient $r = 0.97$ over all genes) between haploid ES cells and control diploid male ES cells (Fig. 2c and Supplementary Fig. 5). In haploid ES cells 279 and 194 genes were more than twofold upregulated or downregulated ($P < 0.05$), respectively (Supplementary Table 2). Among these, 99 X-linked genes were overexpressed and 4 Y-linked genes were lost in haploid ES cells, consistent with different sex chromosome constitutions (Fig. 2d). Thus, haploid ES cells largely maintain a mouse ES cell transcription profile.

This prompted us to investigate the developmental potential of haploid ES cells. For this we introduced a piggyBac transposon vector for expressing green fluorescent protein (GFP) into HAP-2 ES cells. Flow sorting of cells for GFP fluorescence and DNA staining with Hoechst 33342 yielded a haploid ES cell population that expressed GFP at high level, showing that a haploid genome content was maintained during the transfection procedure (Supplementary Fig. 6). GFP-marked haploid ES cells contributed substantially to chimaeric embryos when injected into C57BL/6 blastocysts (Fig. 3a). The great majority of GFP-positive cells extracted from chimaeric embryos had a diploid DNA content (Fig. 3b), indicating that haploid ES cells contributed extensively to development after diploidization. We also obtained two male and two female live-born chimaeras with a substantial contribution from haploid ES cells (Fig. 3c). These mice developed normally with apparent coat colour chimaerism. Similar results were obtained with the HAP-1 and HTG-2 ES cells (Fig. 3d and Supplementary Fig. 7a). Furthermore, the diploid fraction of HAP-2 ES cells at passage 31 could be differentiated into nestin-positive cells after a neural *in vitro* differentiation protocol¹³ (Supplementary Fig. 7b). Taken together, these findings demonstrate that haploid ES cells maintain a wide differentiation potential.

¹Wellcome Trust Centre for Stem Cell Research, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK.

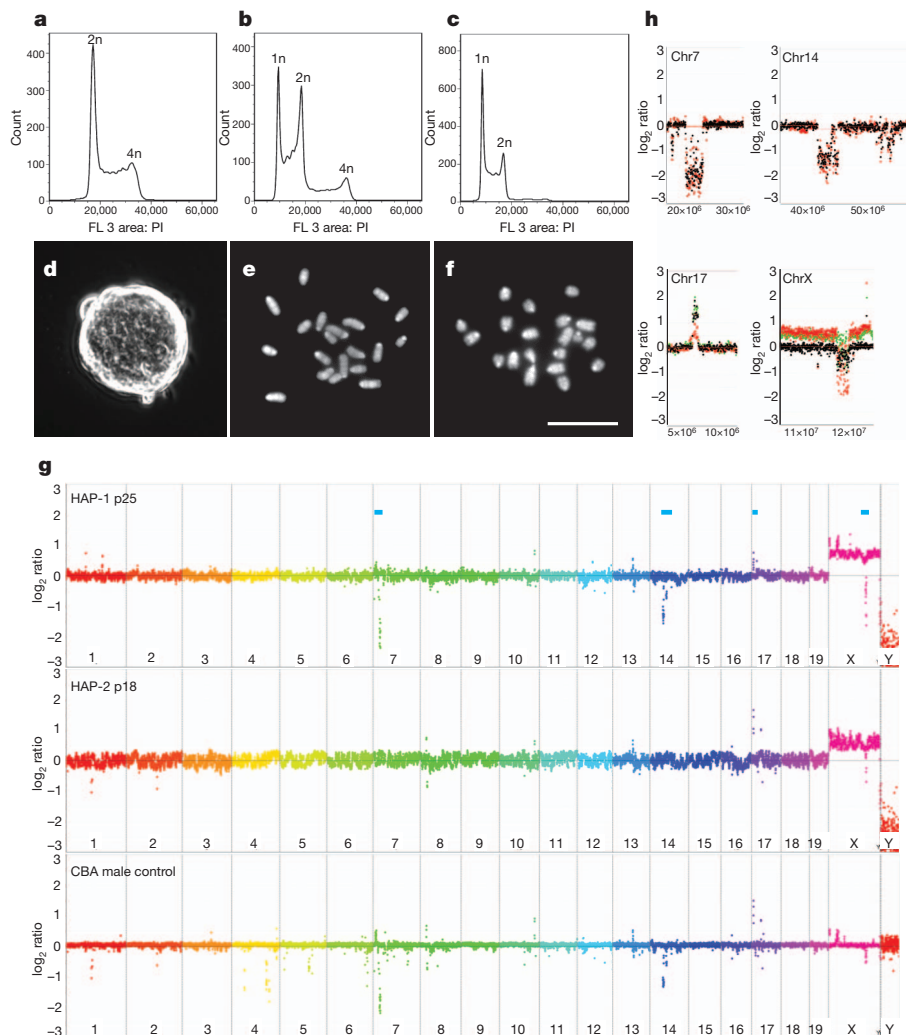


Figure 1 | Derivation of haploid ES cells. **a–c**, Flow analysis of DNA after propidium iodide (PI) staining of diploid control ES cells (**a**), haploid ES cell line HAP-1 at passage 7 (p7) (**b**) and HAP-1 (p11) after sorting at p7 (**c**). The x axis shows fluorescence intensity. **d**, Colony morphology of haploid ES cells (HAP-1). **e, f**, Chromosome spreads of HAP-3 (**e**) and HAP-1 (**f**). Scale bar, 10

μm . **g**, CGH analysis of HAP-1 and HAP-2 ES cells and control male CBA kidney DNA. Relative copy number is plotted at 200-kb resolution using a \log_2 scale. Genomic positions indicated by blue bars (top) are enlarged at 40-kb resolution in **h**. CBA control, black; HAP-1, red; HAP-2, green.

To investigate the utility of haploid ES cells for genetic screening, we performed a pilot screen for mismatch repair genes following a previously published strategy¹⁴. For this, 5×10^6 haploid ES cells were co-transfected with a gene trap piggyBac transposon vector (Supplementary Fig. 8a) and a plasmid for expressing an optimized piggyBac transposase¹⁵. Gene-trap insertions were selected with puromycin. A pool of 1×10^7 cells was then cultured in the presence of 2-amino-6-mercaptopurine (6-TG), which is toxic to mismatch-repair-proficient cells. After 8 days, 20 6-TG-resistant colonies were

isolated and the integration sites were mapped using Splinkerette PCR¹⁶. Of seven clones analysed we identified two independent insertions in *Msh2* and one in *Hprt* (Supplementary Fig. 8b). *Msh2* is a known mismatch repair gene and *Hprt* is required for converting 6-TG into a toxic metabolite¹⁴. Thus, identification of mutations in autosomal genes was possible, indicating a potential for haploid ES cells in forward genetic screening in mammals.

The difficulty in obtaining haploid ES cell lines in previous attempts might be explained by aberrant gene regulation such as aberrant dosage

Table 1 | Derivation of haploid mouse ES cell lines

Experiment no.	Derivation protocol	Names of ES cell lines	Genetic background	Oocytes activated	Number of blastocysts	ES cell lines obtained	ES cell lines with haploid contribution (maximum % haploid before sorting)
1	2i/immunosurgery	HAP-1 to HAP-6*	B6CBAF1	132	30	27	6 (>60%)
2	2i	HAP-7	B6CBAF1	22	10	5	1 (>15%)
3	2i	HTG-1 to HTG-3†	Mixed TG‡	50	32	3	3 (>90%)
4	KSR/immunosurgery	HAP-8 to HAP-13	B6CBAF1	273	48	22	6 (>10%)
5	2i	H129B6-1 to H129B6-5	129B6F1	250	37	8	5 (>10%)
6	2i	HTX-1	Mixed TG§	70	11	1	1 (>40%)
7	2i	H129-1 to H129-3	129Sv	140	13	10	3 (>60%)

KSR, knockout serum replacement.

*Contribution to chimaeric mice was confirmed for the HAP-1 and HAP-2 haploid ES cell lines.

†Contribution to chimaeric mice was confirmed for the HTG-1 haploid ES cell line.

‡Derived from *ROSA26^{nlstTA} LC1 Xist^{2LOX}* homozygous female mice.

§Derived from *ROSA26^{nlstTA} tetOPXist* homozygous female mice²⁶.

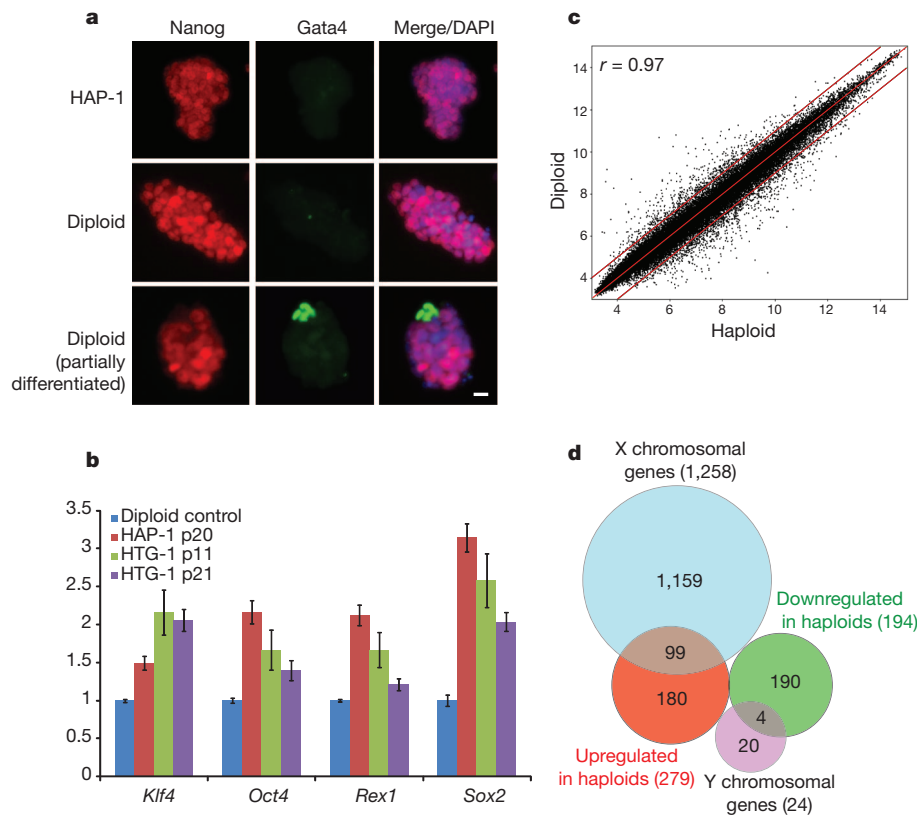
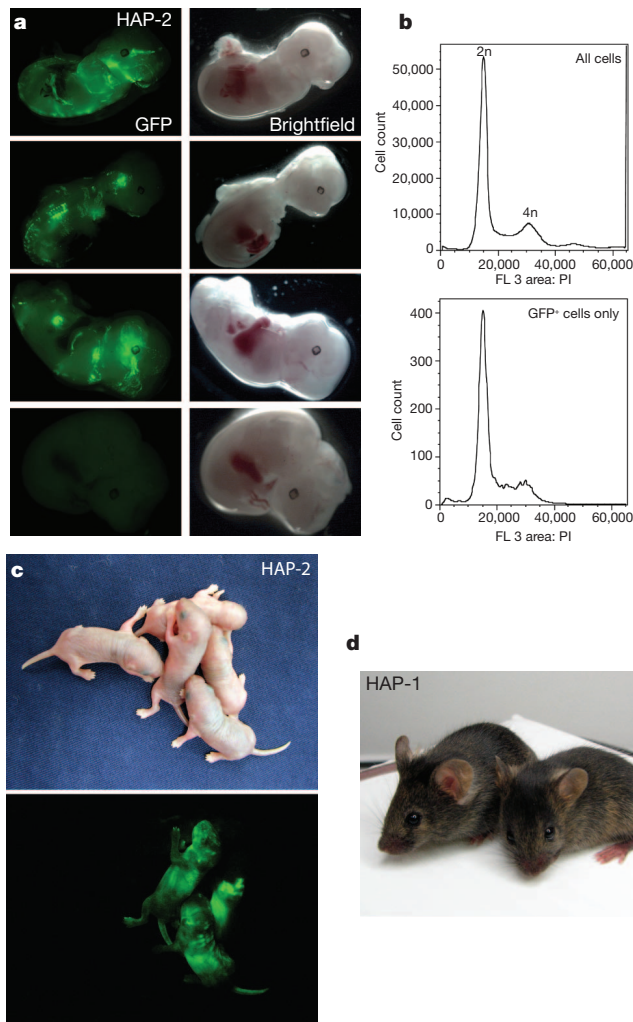


Figure 2 | Expression analysis of haploid ES cells. **a**, Immunofluorescence shows Nanog protein (red) in haploid (HAP-1) and diploid ES cells, and Gata4 (green) in differentiated cells. Scale bar, 10 μ m. **b**, Expression of pluripotency markers in haploid and diploid (set to 1) ES cells by real-time PCR. Error bars represent standard deviation ($n = 3$). **c**, Scatter plot showing \log_2 transformed average expression values from gene expression profiles of three haploid (HAP-1, HAP-2 and HTG-1) and three diploid J1 ES cell lines for 45,001 probe sets (r is the Pearson correlation coefficient; red lines indicate twofold upregulation and downregulation). **d**, Diagram of more than twofold upregulated and downregulated genes in haploid ES cells.



compensation and genomic imprinting. However, diploid ES cells from mouse and human parthenogenetic embryos have been established^{17,18}. Misregulation of X-inactivation has been observed to some extent in haploid mouse embryos⁵ and has also been shown to reduce the efficiency of producing cloned mice¹⁹. Thus, it is conceivable that X-inactivation is initiated aberrantly in haploid embryos during some ES cell derivation procedures. Direct capture of naive pluripotent cells from ICM outgrowths, as accentuated by the use of 2i conditions¹¹, could have contributed to the success of our study.

Previously, near-haploid cells have been observed in human tumours (see ref. 20 for a review), and a near-haploid human-tumour-derived cell line has been described^{21,22}. These tumour cells carry genomic rearrangements and mutations that might stabilize the haploid genome. An interesting aspect of haploid ES cells is their developmental potential. We have observed rapid diploidization when haploid ES cells differentiate. The resulting diploid parthenogenetic cells can contribute to development²³. It is interesting to speculate whether differentiated haploid lineages can be generated, perhaps through suppression of X-inactivation, and whether it is possible to derive haploid human ES cells.

METHODS SUMMARY

For the derivation of haploid ES cells, mouse oocytes were activated in M16 medium as described²⁴. ES cell culture in chemically defined 2i medium has been described previously^{7,8}. Cell sorting for DNA content was performed after staining with 15 μ g ml⁻¹ Hoechst 33342 (Invitrogen) on a MoFlo flow sorter (Beckman Coulter) selecting the haploid 1n peak. For analytic flow profiles, cells were fixed in

Figure 3 | Developmental potential of haploid ES cells. **a**, GFP-marked haploid HAP-2 ES cells (p18) contribute to chimaeric embryos at E12.5. Six out of nine embryos showed GFP contribution. A GFP-negative embryo is shown as a control (bottom). **b**, Representative flow analyses of DNA content of all cells (top) and GFP-positive cells (bottom) extracted from a chimaeric E12.5 embryo are shown. All six embryos gave similar results. **c**, Live-born chimaeric mice were obtained from GFP-marked HAP-2 ES cells. **d**, Chimaeric mice obtained from injection of HAP-1 ES cells into C57BL/6 blastocysts (black) show coat colour contribution from the ES cells (agouti).

ethanol, treated with RNase, and stained with propidium iodide. For karyotype analysis, cells were arrested in metaphase with demecolcine (Sigma). After incubation in hypotonic KCl buffer, cells were fixed in methanol-acetic acid (3:1) and chromosome spreads were prepared and stained with DAPI. RNA was extracted using the RNeasy kit (Qiagen). Transcription profiles were generated using Affymetrix GeneChip 430.2 arrays. Sample preparation, hybridization and basic data analysis were performed by Imagenes. Further analysis was performed using the Genespring GX software (Agilent). For CGH analysis, genomic DNA was isolated from haploid ES cell lines and hybridized to NimbleGen 3x720K whole-genome tiling arrays by Imagenes using C57BL/6 kidney DNA as a reference. For chimera experiments, GFP-labelled HAP-1 (p29), HAP-2 (p18) and HTG-2 (p23) ES cells were injected into C57BL/6 host blastocysts. Live-born chimaeras were analysed for expression of GFP at postnatal day 2. Genetic screening was performed following a previously published strategy²⁵. In brief, HAP-1 ES cells were co-transfected with 2 µg piggyBac transposase expression vector¹⁵ and 1 µg piggyBac gene-trap vector (Supplementary Fig. 8) using Lipofectamine 2000 (Invitrogen). Selection for transposon insertions was performed using 2 µg ml⁻¹ puromycin for 8 days. 1×10^7 puromycin-resistant ES cells were plated in two 15-cm dishes and mutations in mismatch repair genes were selected using 0.3 µg ml⁻¹ 6-TG (Sigma). piggyBac integration sites in seven 6-TG-resistant clones were mapped by Splinkerette PCR¹⁶.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 March; accepted 16 August 2011.

Published online 7 September 2011.

- Otto, S. P. & Jarne, P. Evolution. Haploids—hapless or happening? *Science* **292**, 2441–2443 (2001).
- Wiellette, E. *et al.* Combined haploid and insertional mutation screen in the zebrafish. *Genesis* **40**, 231–240 (2004).
- Yi, M., Hong, N. & Hong, Y. Generation of medaka fish haploid embryonic stem cells. *Science* **326**, 430–433 (2009).
- Kaufman, M. H., Robertson, E. J., Handyside, A. H. & Evans, M. J. Establishment of pluripotential cell lines from haploid mouse embryos. *J. Embryol. Exp. Morphol.* **73**, 249–261 (1983).
- Latham, K. E., Akutsu, H., Patel, B. & Yanagimachi, R. Comparison of gene expression during preimplantation development between diploid and haploid mouse embryos. *Biol. Reprod.* **67**, 386–392 (2002).
- Kaufman, M. H. Chromosome analysis of early postimplantation presumptive haploid parthenogenetic mouse embryos. *J. Embryol. Exp. Morphol.* **45**, 85–91 (1978).
- Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).
- Nichols, J. *et al.* Validated germline-competent embryonic stem cell lines from nonobese diabetic mice. *Nature Med.* **15**, 814–818 (2009).
- Buehr, M. *et al.* Capture of authentic embryonic stem cells from rat blastocysts. *Cell* **135**, 1287–1298 (2008).
- Nichols, J., Silva, J., Roode, M. & Smith, A. Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. *Development* **136**, 3215–3222 (2009).
- Nichols, J. & Smith, A. The origin and identity of embryonic stem cells. *Development* **138**, 3–8 (2011).
- Cutler, G., Marshall, L. A., Chin, N., Baribault, H. & Kassner, P. D. Significant gene content variation characterizes the genomes of inbred mouse strains. *Genome Res.* **17**, 1743–1754 (2007).
- Pollard, S. M., Benchoua, A. & Lowell, S. Neural stem cells, neurons, and glia. *Methods Enzymol.* **418**, 151–169 (2006).
- Li, M. A., Pettitt, S. J., Yusa, K. & Bradley, A. Genome-wide forward genetic screens in mouse ES cells. *Methods Enzymol.* **477**, 217–242 (2010).
- Cadinanos, J. & Bradley, A. Generation of an inducible and optimized piggyBac transposon system. *Nucleic Acids Res.* **35**, e87 (2007).
- Mikkers, H. *et al.* High-throughput retroviral tagging to identify components of specific signaling pathways in cancer. *Nature Genet.* **32**, 153–159 (2002).
- Mai, Q. *et al.* Derivation of human embryonic stem cell lines from parthenogenetic blastocysts. *Cell Res.* **17**, 1008–1019 (2007).
- Revazova, E. S. *et al.* Patient-specific stem cell lines derived from human parthenogenetic blastocysts. *Cloning Stem Cells* **9**, 432–449 (2007).
- Inoue, K. *et al.* Impeding Xist expression from the active X chromosome improves mouse somatic cell nuclear transfer. *Science* **330**, 496–499 (2010).
- Sukov, W. R. *et al.* Nearly identical near-haploid karyotype in a peritoneal mesothelioma and a retroperitoneal malignant peripheral nerve sheath tumor. *Cancer Genet. Cytogenet.* **202**, 123–128 (2010).
- Kotecki, M., Reddy, P. S. & Cochran, B. H. Isolation and characterization of a near-haploid human cell line. *Exp. Cell Res.* **252**, 273–280 (1999).
- Carette, J. E. *et al.* Haploid genetic screens in human cells identify host factors used by pathogens. *Science* **326**, 1231–1235 (2009).
- Jiang, H. *et al.* Activation of paternally expressed imprinted genes in newly derived germline-competent mouse parthenogenetic embryonic stem cell lines. *Cell Res.* **17**, 792–803 (2007).
- Kishigami, S. & Wakayama, T. Efficient strontium-induced activation of mouse oocytes in standard culture media by chelating calcium. *J. Reprod. Dev.* **53**, 1207–1215 (2007).
- Guo, G., Wang, W. & Bradley, A. Mismatch repair genes identified using genetic screens in Bln-deficient embryonic stem cells. *Nature* **429**, 891–895 (2004).
- Savarese, F., Flahndorfer, K., Jaenisch, R., Busslinger, M. & Wutz, A. Hematopoietic precursor cells transiently reestablish permissiveness for X inactivation. *Mol. Cell. Biol.* **26**, 7167–7177 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Smith and J. Nichols for critical discussion; K. Jones for advice on cell culture; S. Dietmann for bioinformatics support; and R. Walker for cell sorting. We would also like to thank B. Mansfield and C.-E. Dumeau for their help, and the BSU team at the centre for maintaining the mouse colony. This work was supported by a Wellcome Trust Senior Research Fellowship to A.W. (grant reference 087530/Z/08/A) and an EMBO Long Term Fellowship to M.L.

Author Contributions M.L. performed the experiments, analysed the data and wrote the manuscript. A.W. performed some experiments, wrote the paper and supervised the study.

Author Information Gene expression and CGH data sets can be accessed as the GEO reference series GSE30879 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30879>). This series includes the GSE30744 (Expression analysis of haploid and diploid ES cells in 2i medium) and the GSE30749 (CGH analysis of haploid ES cells) data sets. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.W. (aw512@cam.ac.uk).

METHODS

Derivation of haploid ES cells. Oocytes were isolated from superovulated females and activated in M16 medium using 5 mM strontium chloride and 2 mM EGTA as described²⁴. Embryos were subsequently cultured in M16 or KSOM medium microdrops covered by mineral oil. Under these conditions around 80% of oocytes reached the 2-cell stage on the next morning. Thereafter development of pre-implantation embryos was variable with a large number of embryos showing unequally sized blastomeres or unusual embryo morphology. Removal of the zona, immunosurgery for removal of the trophectoderm and ES cell derivation were performed as described previously^{7,8}. ES cells were cultured in chemically defined 2i medium plus LIF as described^{7,8} with minor modifications. 2i medium was supplemented with non-essential amino acids and 0.35% BSA fraction V. Culture of ES cells on feeders was performed as previously described²⁷. Knockout serum replacement was obtained from Invitrogen. Cell sorting for DNA content was performed after staining with $15 \mu\text{g ml}^{-1}$ Hoechst 33342 (Invitrogen) on a MoFlo flow sorter (Beckman Coulter). The haploid 1n peak was purified. Diploid cells did arise in cultures to various extents in all ES cell lines. Periodic purification by flow sorting every four to five passages allowed us to maintain cultures containing a great majority of haploid ES cells in all cases. Analytic flow profiles of DNA content were recorded after fixation of the cells in ethanol, RNase digestion and staining with propidium iodide (PI) on a Cyan analyser (Beckman Coulter). For karyotype analysis, cells were arrested in metaphase using demecolcine (Sigma). After incubation in hypotonic KCl buffer, cells were fixed in methanol-acetic acid (3:1) and chromosome spreads were prepared and stained with DAPI. Immunostaining was performed as described²⁸ using Nanog (Abcam; 1:100), Oct4 (Santa Cruz; 1:100), nestin (Developmental Studies Hybridoma bank; 1:30) and Gata4 (Santa Cruz; 1:200) antibodies.

Microarray analysis. RNA from biological triplicates of diploid ES cells and three independently derived haploid ES cells (HAP-1 p21, HAP-2 p24, HTG-1 p11) was extracted using the RNeasy kit (Qiagen). Gene expression analysis on Affymetrix GeneChip 430 2.0 arrays was performed by Imagenes. Additional gene expression profiles of neural progenitors, mesodermal progenitors and mouse embryonic fibroblasts (MEFs) were obtained from a previously published data set (GEO accession number GSE12982 (ref. 29)). The data were analysed using Genespring GX software (Agilent Technologies). Data were normalized using the RMA algorithm. Lists showing differentially regulated genes (>2 -fold change; $P < 0.05$) are provided in Supplementary Table 2. P values were established by an unpaired t -test followed by FDR adjustment by the Benjamini Hochberg method. Hierarchical clustering was performed based on the Euclidean distances and complete linkage analysis. The relatedness of transcription profiles was determined by calculating the Pearson correlation coefficient (r). DNA samples for CGH experiments were extracted and sent to Imagenes for CGH analysis using NimbleGen 3x720K mouse whole-genome tiling arrays with an average probe spacing of 3.5 kb. Adult male C57BL/6 kidney DNA was used as a reference. A genomic overview of these analyses is presented in Fig. 1g and Supplementary Fig. 3 at

200-kb resolution and selected magnified regions at 40-kb resolution. The complete data set at 40-kb resolution is included in Supplementary Fig. 4.

Accession of data sets. Gene expression and CGH data sets can be accessed as the GEO reference series GSE30879 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30879>). This series includes the GSE30744 (Expression analysis of haploid and diploid ES cells in 2i medium) and the GSE30749 (CGH analysis of haploid ES cells) data sets.

Quantitative gene expression analysis. RNA was extracted using the RNeasy kit (Qiagen) and converted into cDNA using the Quantitect reverse transcription kit (Qiagen). Real-time PCR was performed on a StepOnePlus machine (Applied Biosystems) using the Fast Sybr green master mix (Applied Biosystems) and previously published primers²⁸. The $\Delta\Delta C_t$ method was used for quantification of gene expression. Expression levels were normalized to L32 ribosomal protein mRNA and values in diploid control ES cells were set to 1.

Embryo analysis. Haploid ES cells were co-transfected with a piggyBac vector carrying a CAG-GFP-IRES-hygro transgene and a piggyback transposase expression plasmid. Stable integrants were selected using $150 \mu\text{g ml}^{-1}$ hygromycin for 7 days. The haploid fraction of HAP-1 (p29), HAP-2 (p18) and HTG-2 (p23), GFP-positive cells were purified by flow sorting (Supplementary Fig. 6). GFP-labelled ES cells were expanded and injected into C57BL/6 host blastocysts which were transferred to recipient females. Embryos were analysed at E9.5 and E12.5. Dissociation to single cells was performed by incubation in 0.25% trypsin/EDTA for 15 min. Before PI staining, cells were fixed in 4% PFA and permeabilized in PBS/0.25% Triton X-100. Live-born chimaeras were analysed at postnatal day 2 (P2) for expression of GFP using ultraviolet illumination. Images were obtained using a Canon Powershot S5 IS camera with a FHS/EF-3GY2 filter (BLS). All mouse experiments were conducted in accordance with institutional guidelines of the University of Cambridge. All necessary UK home office licenses were in place.

Gene-trap screen. The screen was performed based on a previously published protocol²⁵. 5×10^6 HAP-1 ES cells were co-transfected with $2 \mu\text{g}$ piggyBac transposase plasmid¹⁵ and $1 \mu\text{g}$ piggyBac gene-trap vector (Supplementary Fig. 8a) using Lipofectamine 2000. piggyBac insertions into expressed genes were selected with $2 \mu\text{g ml}^{-1}$ puromycin for 8 days. 1×10^7 ES cells corresponding to approximately 5,000 puromycin-resistant colonies were then plated onto two 15 cm dishes. Selection for mismatch-deficient integrants was performed using $0.3 \mu\text{g ml}^{-1}$ 6-TG (Sigma). Twenty colonies were picked and piggyBac integration sites of seven clones were identified by Splinkerette PCR and mapped using iMapper³⁰.

27. Wutz, A. & Jaenisch, R. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol. Cell* **5**, 695–705 (2000).
28. Leeb, M. *et al.* Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev.* **24**, 265–276 (2010).
29. Shen, X. *et al.* EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol. Cell* **32**, 491–502 (2008).
30. Kong, J., Zhu, F., Stalker, J. & Adams, D. J. iMapper: a web application for the automated analysis and mapping of insertional mutagenesis sequence data against Ensembl genomes. *Bioinformatics* **24**, 2923–2925 (2008).

RNAi promotes heterochromatic silencing through replication-coupled release of RNA Pol II

Mikel Zaratiegui^{1†}, Stephane E. Castel^{1,2}, Danielle V. Irvine^{1†}, Anna Kloc^{1†}, Jie Ren¹, Fei Li^{3†}, Elisa de Castro⁴, Laura Marín⁴, An-Yun Chang^{1,5}, Derek Goto^{1†}, W. Zacheus Cande³, Francisco Antequera⁴, Benoit Arcangeli^{1,6} & Robert A. Martienssen^{1,2}

Heterochromatin comprises tightly compacted repetitive regions of eukaryotic chromosomes. The inheritance of heterochromatin through mitosis requires RNA interference (RNAi), which guides histone modification¹ during the DNA replication phase of the cell cycle². Here we show that the alternating arrangement of origins of replication and non-coding RNA in pericentromeric heterochromatin results in competition between transcription and replication in *Schizosaccharomyces pombe*. Co-transcriptional RNAi releases RNA polymerase II (Pol II), allowing completion of DNA replication by the leading strand DNA polymerase, and associated histone modifying enzymes³ that spread heterochromatin with the replication fork. In the absence of RNAi, stalled forks are repaired by homologous recombination without histone modification.

In fission yeast, the Rik1/CLRC (recombination in K, cryptic locus regulator) complex silences heterochromatin via Clr4 and Lid2, which methylate histone H3 lysine 9 (H3K9) and demethylate histone H3 lysine 4 (H3K4), respectively². This complex is recruited in part by RNA interference, which processes non-coding transcripts found in the pericentromeric heterochromatin^{1,4}. Interactions between the RNA-induced transcriptional silencing (RITS) complex and CLRC have recently been found^{5,6}, but spreading of the Rik1 complex into reporter genes depends on the catalytic activity of RNAi, and the mechanism remains unknown⁷. Recently, we found that Cdc20 and Mms19 interact with Rik1 and are required for histone modification³. Cdc20 is the catalytic subunit of the leading strand DNA polymerase Pol ε, whereas Mms19 is a regulatory subunit of the Pol II transcription factor TFIIF. Both proteins participate in transcription coupled nucleotide excision repair (TC-NER) which depends on damage-stalled Pol II to detect structural lesions in the DNA which are repaired by the Pol ε after Pol II release⁸.

The pericentromeric heterochromatin of fission yeast comprises outermost (*otr*) repeats called *dg* (5 kilobases) and *dh* (1–6 kb), flanked by innermost (*imr*) repeats (~6 kb) containing clusters of transfer RNA genes (Fig. 1a). Histone H3 lysine-9 methylation is associated with *dg* and *dh* repeats (Fig. 1b), but ends abruptly at the tRNA clusters, and so is confined to heterochromatin⁹. The *dg* and *dh* repeats are transcribed by RNA polymerase II¹⁰, and processed into siRNA clusters up to 4.5 kb in length (Fig. 1b). To investigate the extent of siRNA precursor transcripts, we first cloned and sequenced *dh* and *dg* repeat complementary DNA from *dcr1Δ* mutants (Fig. 1c). Polyadenylation sites were then identified using rapid amplification of cDNA ends-PCR (RACE-PCR; Methods), and sequencing showed they were located within the clusters of short interfering RNA (siRNA; Fig. 1c). In previous studies of *dcr1Δ* mutants, Pol II enrichment was detected by chromatin immunoprecipitation (ChIP)¹¹, whereas transcriptional run-on (TRO) analysis indicated over-accumulation of forward (but not reverse) transcripts¹. We found that these Pol II ChIP (cen-*dg*) and TRO probes lie downstream of 'forward' polyA sites (Fig. 1d),

indicating inefficient termination and Pol II readthrough in the absence of RNAi. To confirm readthrough, northern blots of polyadenylated and total RNA from *dcr1Δ*, *ago1Δ* and *rdp1Δ* mutants were probed with strand-specific probes. Transcripts corresponding to full-length *dh* (1.3 kb) and *dg* (1.3–2.3 kb) cDNA clones were enriched in polyA+ RNA, as expected, but much longer readthrough transcripts up to 4.5 kb could also be detected (Supplementary Fig. 1), indicating that polyadenylation was highly inefficient at these internal sites.

Inefficient polyadenylation is a strong indication of failure to release RNA polymerase II¹², and we proposed that slicing⁷ and dicing¹³ of nascent transcripts via RNAi promotes 3'–5' degradation by the exosome⁷ and release of RNA polymerase II from the 3' end¹². The exosome is required for silencing, consistent with this idea^{14,15}. To examine Pol II release, we performed ChIP-seq with Pol II antibodies, and found peaks of both poised (S5 phosphorylated) and elongating forms (S2 phosphorylated) of Pol II in *dcr1Δ* mutants that corresponded to the polyadenylation sites on each strand (Fig. 1b, c). Peaks of siRNA accumulation mapped just downstream. Thus siRNA in wild-type cells accumulated where Pol II was released (Fig. 1b).

siRNA accumulate during S phase² and we found that siRNA clusters ended abruptly at the replication origin homology regions contained within each repeat¹⁶ (Fig. 1b). To assess the influence of DNA replication on Pol II accumulation we blocked replication in high concentrations of hydroxyurea and performed ChIP-seq using Pol II antibodies. In arrested *dcr1Δ* mutants, Pol II accumulated throughout the *otr* repeats, but in dividing *dcr1Δ* cells, Pol II accumulation was absent from replication origins (Fig. 2a and data not shown). To test if Pol II was expelled by replication fork progression (Fig. 2b), hydroxyurea-arrested *dcr1Δ* cells were released into the cell cycle (Fig. 2c). As predicted, accumulation at replication origins was quickly lost, and Pol II was only found between origins, closer to promoters¹⁰, in each subsequent S phase.

Failure to release RNA polymerase II during S phase is a strong and robust signal for DNA damage⁸. To monitor DNA repair, the hydroxyurea-arrested cells contained a Rad22 protein fused with yellow fluorescent protein (Rad22-YFP). Rad22 (Rad52 in budding yeast) is essential for homologous recombination and is associated with single-stranded DNA ends early during DNA repair. Chromatin immunoprecipitation showed that Rad22^{Rad52} was weakly associated with heterochromatic origins in wild-type cells arrested with hydroxyurea, but quickly declined following release (Fig. 2c). In *dcr1Δ* mutants, on the other hand, Rad22^{Rad52} peaked early in each successive S phase, indicating engagement of the repair machinery during heterochromatin replication¹⁷. To exclude the impact of hydroxyurea arrest on DNA damage, we also examined Rad22-YFP accumulation in untreated wild-type and *dcr1Δ* mutant cells by fluorescence microscopy (Supplementary Fig. 2). The results were consistent with chromatin immunoprecipitation, in that six times as many *dcr1Δ* than

¹Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. ²Watson School of Biological Sciences, Cold Spring Harbor Laboratory, New York 11724, USA. ³Molecular and Cellular Biology, University of California Berkeley, Berkeley 94720, USA. ⁴Instituto de Biología Funcional y Genómica. CSIC/Universidad de Salamanca, Salamanca 37007, Spain. ⁵Molecular and Cell Biology program, Stony Brook University, Stony Brook, New York 11794, USA. ⁶Institut Pasteur, Paris, France. [†]Present addresses: Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey 08854, USA (M.Z.); Yale Stem Cell Center, Yale University, New Haven, Connecticut 06520, USA (A.K.); Creative Research Initiative Sousei, Hokkaido University, 001-0021 Sapporo, Japan (D.G.); Department of Biology, New York University, New York, New York 10003, USA (F.L.); Murdoch Children's Research Institute, University of Melbourne, Melbourne, Victoria 3052, Australia (D.V.I.).

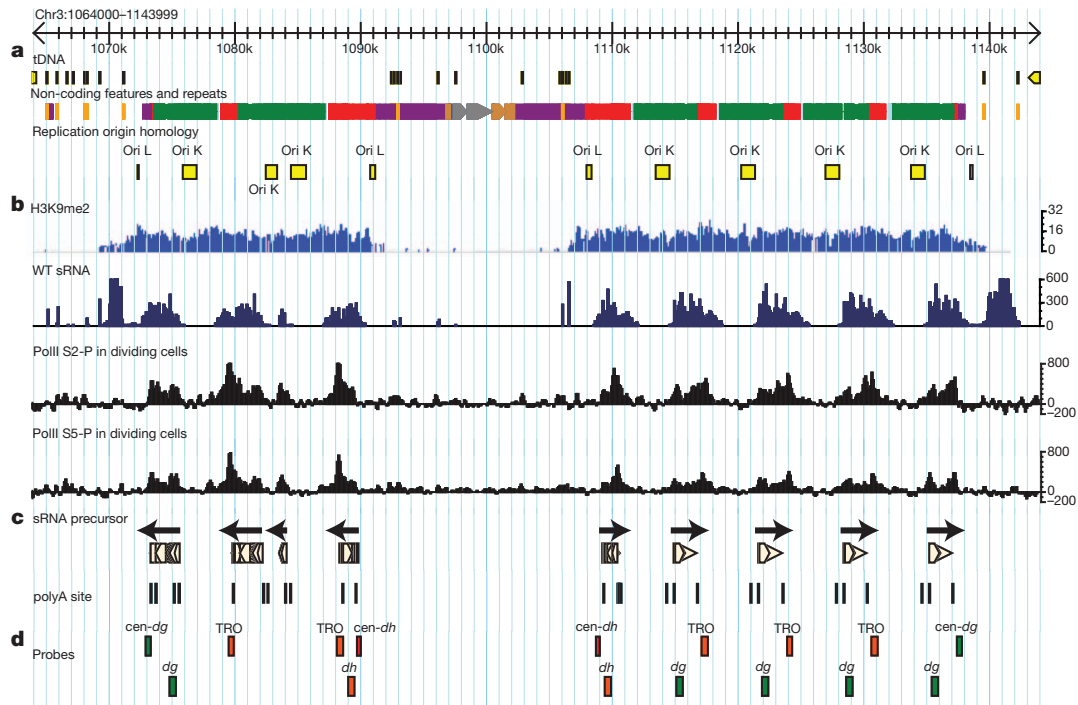


Figure 1 | Transcription and replication of pericentromeric heterochromatin in fission yeast. **a**, Pericentromeric heterochromatin on centromere 3. *dh* (red), *dg* (green) and *imr* (magenta) repeats are shown, bordered by tRNA genes (brown). Replication origins (yellow) are found in each repeat. **b**, Tiling microarrays of K9me2 ChIP (light blue) and clusters of small RNA sequences (dark blue) from wild-type cells. ChIP-seq reads corresponding to poised (S5-P) and elongating (S2-P) RNA polymerase II

enriched in *dcr1Δ* cells relative to wild-type (WT) cells are in black. **c**, cDNA clones (beige) from *dcr1Δ* cells. PolyA sites are indicated as vertical lines and correspond to peaks of Pol II. Arrows indicate the direction of 'Forward' transcription. **d**, Alignment of probes used in previous studies indicates that regions enriched for Pol II¹¹ (*cen-dg*) and transcriptional run-on probes¹ (TRO) lie downstream of forward orientation polyA sites.

wild-type cells had Rad22^{Rad52} foci during septation (early S phase). Therefore, Dcr1 activity prevents DNA damage and the engagement of homologous recombination at the centromere.

We performed genetic tests to determine the role of RNAi in preventing DNA damage during S phase. DNA damage during replication can be rescued by homologous recombination repair, and we found that double mutants in the RecA homologue *rhp51^{rad51}* and *dcr1Δ* or *ago1Δ* were inviable or formed microcolonies (Fig. 2d). A similar requirement for Rhp51^{Rad51} has been demonstrated for convergent stalled replication forks¹⁸, which are protected from collapse in fission yeast by a stable replication-pausing complex comprising Swi1/Swi3 and Mrc1 (mediator of replication checkpoint 1)¹⁹. Low concentrations of hydroxyurea stall replication forks, and we found that whereas *dcr1Δ*, *ago1Δ* and *rdp1Δ* cells were insensitive, double mutants with *swi3Δ* or *mrc1Δ* were very sensitive to low concentrations of hydroxyurea (Supplementary Fig. 3). Similar results were obtained with camptothecin which causes arrest during S phase when the replication fork encounters the camptothecin-topoisomerase I complex. In genome-wide epistasis tests, mutants in more than 30 genes, mostly encoding proteins involved in DNA repair and histone modification, interacted significantly with both *mrc1* and *dcr1*, forming a striking genetic network (Supplementary Table 1). This indicates that loss of Dcr1 activity engages replication fork protection.

To assess fork integrity, we examined replication of the repeats by two-dimensional gel electrophoresis using probes from the *ura4* transgene, which was inserted into a passively replicated *dg* repeat on chromosome 1 (Fig. 2b). In wild-type cells, we detected strong X intermediates, indicative of joint molecules, as well as the expected fork or Y molecules (Fig. 3a). Similar X-DNA sister chromatid junctions arise at origins²⁰ but also at stalled replication forks²¹. These X-molecules were unaffected in *dcr1Δ* (Fig. 3b) but reduced in *mms19Δ*, in *swi6Δ* and especially in *clr4Δ* cells (Fig. 3c–e). Both Mms19 and Clr4 interact with Rik1, and Mms19 participates in

transcription initiation³. Swi6 on the other hand is required to initiate replication within heterochromatic repeats¹⁷, and recruitment depends on Clr4. Thus simultaneous replication and transcription of heterochromatic repeats promote local replication fork stalling.

In wild-type cells (Fig. 4a), modified histones recruit Swi6 and the Rik1 complex via chromo- and other domains. Swi6 promotes early replication, and the Rik1 complex interacts with DNA polymerase ϵ , which allows spreading of histone modification along with fork progression³. Flanking tRNA genes (Fig. 1a) pause replication²², preventing further spreading into neighbouring euchromatin^{9,23}. Transcription during S phase stalls the replication fork, accounting for interactions between the replication and transcription machineries³, but RNAi releases Pol II, allowing replication to proceed. In the absence of RNAi (Fig. 4b), Pol II remains stalled at replication forks and signals DNA repair by homologous recombination, which restarts blocked forks²⁴. The Rik1 complex is lost along with the replisome, preventing spreading of heterochromatin into reporter genes, which lose H3K9 methylation entirely. Recombination also removes modified histones from at least one of the two daughter chromatids²⁵ reducing, but not eliminating, methylation of the repeats as previously observed⁷.

We tested this model in several ways. First, we predicted that the interaction between the Rik1 complex and Pol ϵ should depend on RNAi, and we found that co-immunoprecipitation of Cdc20/Pol ϵ with Dos2/Clr7 was reduced in *dcr1Δ* cells, along with H3K9me2 (Supplementary Fig. 4). Second, we observed that mutants in the cyclin-dependent Pol II C-terminal domain kinase *cdk9* display slow growth and loss of pericentromeric silencing and siRNA (Supplementary Fig. 5). Cdk9 is a central regulator of transcription elongation that links cell-cycle-regulated pre-messenger RNA processing, co-transcriptional histone methylation and DNA damage²⁶. Finally, Clr4 has recently been found to have additional roles in recruiting the RITS complex to accessory Pol II factors²⁷, providing a potential mechanism for Pol II release by RNAi. We found long transcripts

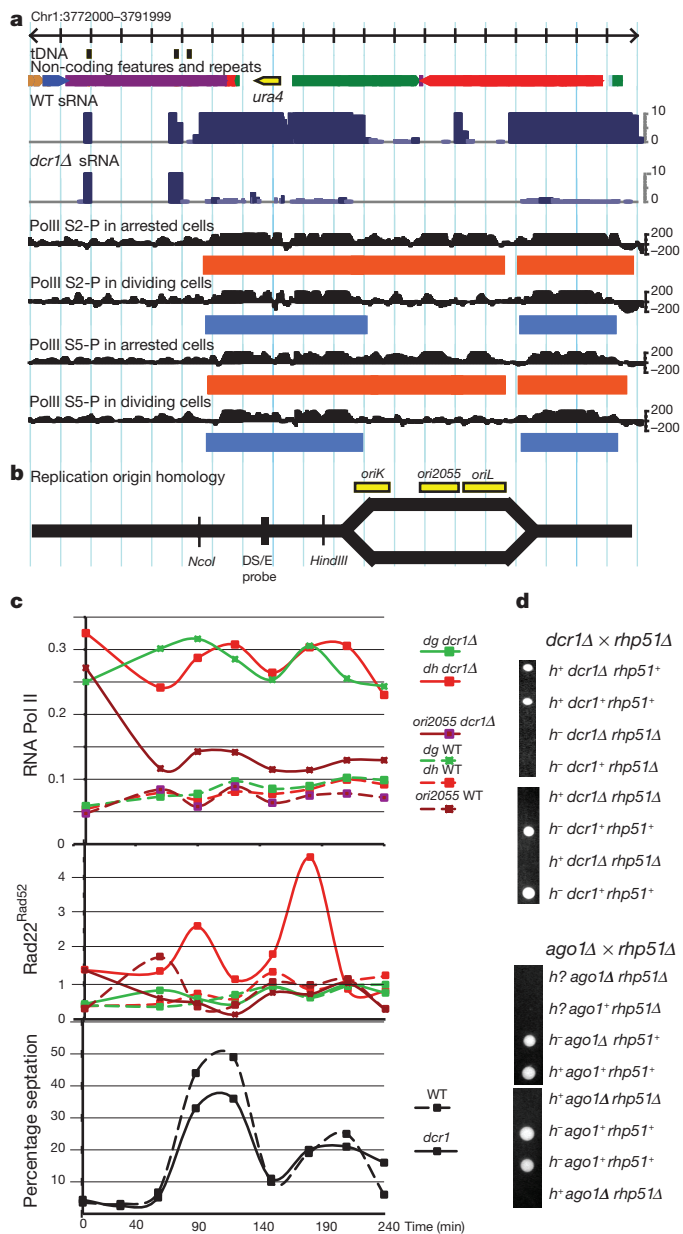


Figure 2 | RNA interference and DNA replication restrict RNA polymerase II accumulation and prevent DNA damage. **a**, Small RNA (blue) and Pol II ChIP-seq reads (black) and regions of significant enrichment (blue and red rectangles) from wild type and *dcr1Δ* on the right arm of centromere 1. **b**, A replication bubble is shown, initiated at one of the three origin homology regions at centromere 1 (yellow boxes). **c**, Chromatin immunoprecipitation for RNA Pol II and Rad22^{Rad52} from hydroxyurea-arrested and released wild-type (dashed lines) and *dcr1Δ* (solid lines). Cell cycle progression after release from hydroxyurea block is monitored by septation index, which peaks coincident with S phase. **d**, Representative parental and non-parental di-type tetrads from crosses between *rhp51Δ* cells, defective in homologous recombination, and *dcr1Δ* or *ago1Δ*.

indicative of strong transcriptional readthrough in *clr4Δ* mutant cells consistent with this model (Supplementary Fig. 1).

In the budding yeast *Saccharomyces cerevisiae*, the Dicer-related RNase III Rnt1 releases Pol II during transcription termination²⁸ whereas in *Escherichia coli*, failure of transcription termination stalls replication forks and triggers recombination²⁹, providing a precedent for the mechanism we propose. According to this mechanism, transcription during S phase triggers histone modification, so long as RNA polymerase is released by RNAi, and not by homologous

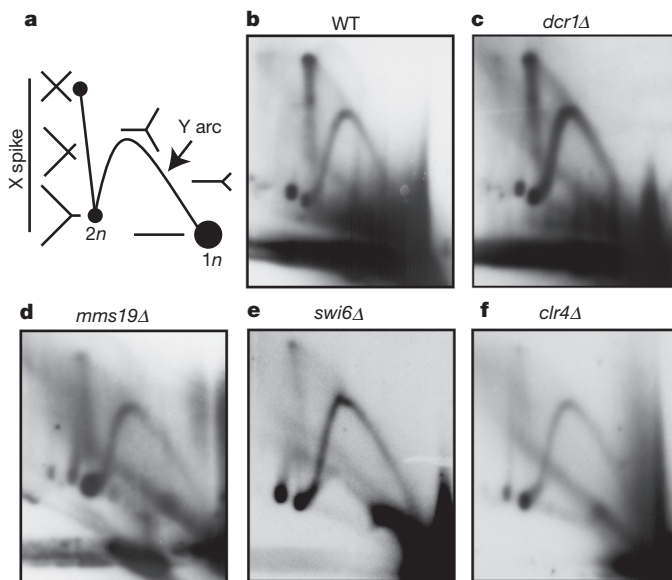


Figure 3 | Replication fork stalling during heterochromatin replication. Replication intermediates in wild-type and mutant cells resolved by 2D gel electrophoresis and probed with the unique DS/E probe from the *ura4* transgene within the *dg* repeat on chromosome 1 (Fig. 2a). **a**, Schematic of replication intermediates in 2D gels indicates joint molecules (X), and forks (Y). **b–f**, Junction molecules indicate fork stalling in WT (**b**) and *dcrl1Δ* mutant cells (**c**), and are reduced in *mms19Δ* (**d**), *swi6Δ* (**e**) and *clr4Δ* (**f**).

recombination repair. In plants, fungi and invertebrates, heterochromatic silencing may involve similar mechanisms (Supplementary Table 1), whereas in mammals, both X inactivation and imprinting require transcription of non-coding RNA in dividing cells³⁰. In each case, release of Pol II during S phase, by RNAi or by other means, could

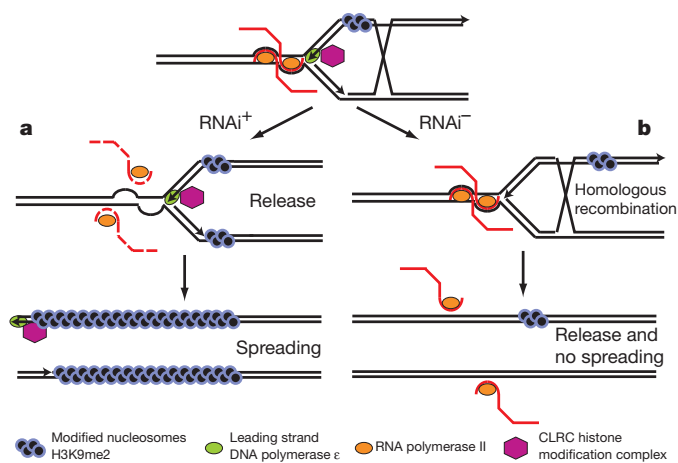


Figure 4 | Replication-coupled transcriptional silencing through histone modification and RNAi. **a**, The Rik1 complex (red hexagon) is recruited to heterochromatic replication forks by interactions with methylated histone H3K9me2 and with the leading strand DNA polymerase (Pol ϵ , green). Swi6 induces origin firing, but collision with RNA polymerase II (orange) stalls replication forks. RNAi releases Pol II by processing of pre-siRNA transcripts (red lines), allowing leading strand DNA polymerase to complete DNA replication and the associated Rik1 histone modification complex (red hexagon) to spread histone modification (black circles). **b**, In the absence of RNAi, origins fire but Pol II is not released, stalling replication forks. Stalled Pol II signals repair via homologous recombination instead. Recombination could in principle occur with sister chromatids (shown here) or with other copies of the same repeat (not shown). DNA polymerase ϵ and the associated Rik1 complex are lost along with the replisome, and fail to spread histone modification into neighbouring reporter genes.

allow fork restart and spreading of histone modification in a similar way.

METHODS SUMMARY

Non-coding transcripts were cloned from a cDNA phage library by hybridization to *dh* and *dg* consensus probes. Cloning and high throughput sequencing of sRNA was performed using an Illumina Genome Analyzer according to manufacturer's instructions. Two-dimensional gel electrophoresis of replication intermediates from steady state cultures was performed with probes to the *otr1::ura4⁺* insertion. For ChIP experiments, cultures were arrested in 15 mM hydroxyurea for 4.5 h, released and collected at indicated times, to be crosslinked and processed for chromatin immunoprecipitation.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 January; accepted 25 August 2011.

Published online 16 October 2011.

- Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).
- Kloc, A., Zaratiegui, M., Nora, E. & Martienssen, R. RNA interference guides histone modification during the S phase of chromosomal replication. *Curr. Biol.* **18**, 490–495 (2008).
- Li, F., Martienssen, R. & Cande, W. Z. Coordination of DNA replication and histone modification by the Rik1–Dos2 complex. *Nature* **475**, 244–248 (2011).
- Verdel, A. *et al.* RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672–676 (2004).
- Motamedi, M. R. *et al.* HP1 proteins form distinct complexes and mediate heterochromatic gene silencing by nonoverlapping mechanisms. *Mol. Cell* **32**, 778–790 (2008).
- Bayne, E. H. *et al.* Stc1: a critical link between RNAi and chromatin modification required for heterochromatin integrity. *Cell* **140**, 666–677 (2010).
- Irvine, D. V. *et al.* Argonaute slicing is required for heterochromatic silencing and spreading. *Science* **313**, 1134–1137 (2006).
- Svejstrup, J. Q. The interface between transcription and mechanisms maintaining genome integrity. *Trends Biochem. Sci.* **35**, 333–338 (2010).
- Cam, H. P. *et al.* Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nature Genet.* **37**, 809–819 (2005).
- Djupedal, I. *et al.* RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. *Genes Dev.* **19**, 2301–2306 (2005).
- Bühler, M., Verdel, A. & Moazed, D. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* **125**, 873–886 (2006).
- Rosonina, E., Kaneko, S. & Manley, J. L. Terminating the transcript: breaking up is hard to do. *Genes Dev.* **20**, 1050–1056 (2006).
- Djupedal, I. *et al.* Analysis of small RNA in fission yeast; centromeric siRNAs are potentially generated through a structured RNA. *EMBO J.* **28**, 3832–3844 (2009).
- Bühler, M., Haas, W., Gygi, S. P. & Moazed, D. RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. *Cell* **129**, 707–721 (2007).
- Murakami, H. *et al.* Ribonuclease activity of Dis3 is required for mitotic progression and provides a possible link between heterochromatin and kinetochore function. *PLoS ONE* **2**, e317 (2007).
- Smith, J. G. *et al.* Replication of centromere II of *Schizosaccharomyces pombe*. *Mol. Cell Biol.* **15**, 5165–5172 (1995).
- Hayashi, M. T., Takahashi, T. S., Nakagawa, T., Nakayama, J. & Masukata, H. The heterochromatin protein Swi6/HP1 activates replication origins at the pericentromeric region and silent mating-type locus. *Nature Cell Biol.* **11**, 357–362 (2009).
- Lambert, S., Watson, A., Sheedy, D. M., Martin, B. & Carr, A. M. Gross chromosomal rearrangements and elevated recombination at an inducible site-specific replication fork barrier. *Cell* **121**, 689–702 (2005).
- Shimmoto, M. *et al.* Interactions between Swi1–Swi3, Mrc1 and S phase kinase, Hsk1 may regulate cellular responses to stalled replication forks in fission yeast. *Genes Cells* **14**, 669–682 (2009).
- Segurado, M., Gomez, M. & Antequera, F. Increased recombination intermediates and homologous integration hot spots at DNA replication origins. *Mol. Cell* **10**, 907–916 (2002).
- Minca, E. C. & Kowalski, D. Multiple Rad5 activities mediate sister chromatid recombination to bypass DNA damage at stalled replication forks. *Mol. Cell* **38**, 649–661 (2010).
- Deshpande, A. M. & Newlon, C. S. DNA replication fork pause sites dependent on transcription. *Science* **272**, 1030–1033 (1996).
- Scott, K. C., Merrett, S. L. & Willard, H. F. A heterochromatin barrier partitions the fission yeast centromere into discrete chromatin domains. *Curr. Biol.* **16**, 119–129 (2006).
- Lambert, S. *et al.* Homologous recombination restarts blocked replication forks at the expense of genome rearrangements by template exchange. *Mol. Cell* **39**, 346–359 (2010).
- Groth, A., Rocha, W., Verreault, A. & Almouzni, G. Chromatin challenges during DNA replication and repair. *Cell* **128**, 721–733 (2007).
- Pirngruber, J., Shchebet, A. & Johnsen, S. A. Insights into the function of the human P-TEFb component CDK9 in the regulation of chromatin modifications and co-transcriptional mRNA processing. *Cell Cycle* **8**, 3636–3642 (2009).
- Zhang, K. *et al.* Ctr4/Suv39 and RNA quality control factors cooperate to trigger RNAi and suppress antisense RNA. *Science* **331**, 1624–1627 (2011).
- Ghazal, G. *et al.* Yeast RNase III triggers polyadenylation-independent transcription termination. *Mol. Cell* **36**, 99–109 (2009).
- Washburn, R. S. & Gottesman, M. E. Transcription termination maintains chromosome integrity. *Proc. Natl Acad. Sci. USA* **108**, 792–797 (2011).
- Pauler, F. M., Koerner, M. V. & Barlow, D. P. Silencing by imprinted noncoding RNAs: is transcription the answer? *Trends Genet.* **23**, 284–292 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Roh and T. Volpe for isolating cDNA clones. D.V.I. was supported by a NHMRC CJ Martin Postdoctoral Research Fellowship. M.Z. was supported by a fellowship from the Spanish Ministry of Science. This work was supported by grants BFU2008-01919 and Consolider-Ingenio CSD2007-00015 from the Spanish Ministry of Science and Innovation to F.A., and NIH R01 GM076396 to W.Z.C. and R.A.M.

Author contributions S.C., D.V.I., A.K., J.R. contributed equally to this work and are listed in alphabetical order. M.Z., S.C., D.V.I., A.K., J.R., F.L., E.d.C., L.M., A.-Y.C. and D.G. performed experiments, and S.C. analysed the data. W.Z.C., F.A., B.A. and R.A.M. designed experiments and R.A.M. and M.Z. wrote the manuscript.

Author Information Genomics data and analysis are available from the Gene Expression Omnibus accession number GSE30837. Individual cDNA sequences are available from GenBank with accession numbers JN388396–JN388565. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.A.M. (martiens@cshl.edu).

METHODS

Yeast strains and methods. *S. pombe* strains used in this study are listed in Supplementary Table 2. Standard media and genetic protocols for fission yeast were used³¹. Crosses between *dcr1*, *ago1* and *rhp51* deletion mutant strains were performed with strains of h^+ , h^- and *smt-0* mating types, with comparable results. Lethality or sickness was observed in 90% of the double mutants as well as many single mutants, suggesting epigenetic effects inherited through meiosis.

cDNA analysis. Total RNA was extracted from cells growing in YES by the hot phenol method. In brief, mid-log phase (0.5 attenuation at 600 nm) yeast cultures were incubated in SDS-acetate buffer (50 mM sodium acetate, 10 mM EDTA, 1% SDS pH 5), with an equal volume of phenol pH 4.3 at 65 °C for 15 min, vortexing every 2 min for 30 s. After centrifugation, the aqueous phase was extracted once with acid phenol, once with acid phenol:chloroform 1:1 and precipitated by addition of 1/10 volume of 3 M sodium acetate pH 8 and 2 volumes of ethanol, resuspended and treated with a DNA-free kit (Ambion). cDNA was reverse-transcribed from total RNA using oligodT adapters, and cloned into phage vectors as recommended by the supplier (Stratagene). More than 60 clones were purified by blot hybridization to *dg* and *dh* probes, and sequenced from both ends. Polyadenylation sites were identified by a string of T residues longer than the adaptor followed by matches to the *S. pombe* consensus sequence. Alignment to the genome sequence was possible because of single nucleotide polymorphisms between copies of the repeats, with the exception of chromosome 3, where the tandemly arranged *dg* and *dh* repeats are identical. Sequences containing a long stretch of A residues immediately downstream of known centromeric sequence were considered to represent the polyadenylated 3' ends of centromeric transcripts. Polyadenylation sites of centromeric cDNA sequences from wild-type and *rdl1Δ* strains were determined using the ligation-mediated polyA test (LM-PAT) assay as previously described³² with the following message-specific primers: dgIIIrevPolC, dgIIIrevPolD, dgIIIforPolC, dhlforPolA and dhlrevPolA. The PCR products amplified from LM-PAT cDNA were separated on an agarose gel, individually extracted and cloned into the pCR2.1-TOPO vector (Invitrogen) according to the manufacturer's instructions. Cloned products were then sequenced in both directions. See oligonucleotide list for details (Supplementary Table 3).

Northern blot analysis. Total RNA was extracted from cells growing in YES by the hot phenol method. PolyA⁺ RNA was isolated from wild-type and mutant strains using Dynabeads Oligo(dT) (Invitrogen). Northern blot analysis was performed using NorthernMax-Gly (Ambion). Blots were probed for the *dh* transcripts (p30_F_T7 and p30_R_T3), *dg* transcripts (p33_F_T7 and p33_R_T3) and actin (act1F and act1R_T7). See oligonucleotide list (Supplementary Table 3) for details.

Small RNA library sequencing. Total RNA was extracted from cells growing at exponential phase in YES using hot phenol extraction. Small RNAs were isolated using flashPAGE followed by gel isolation. Small RNA libraries for Solexa/Illumina sequencing were prepared according to the manufacturers' instructions. Libraries were sequenced on an Illumina Genome Analyzer. The resulting sequence reads, including quality scores, were aligned to the *S. pombe* reference sequence with MAQ (<http://maq.sourceforge.net/>) using the 'map' function.

Two-dimensional gel electrophoresis. Two-dimensional gel electrophoresis of centromeric DNA was performed as previously described³. DNA pellets were digested with NcoI/HindIII to generate restriction fragments of 3.7 kb for origin 2055 and 4.0 kb for the *otr::ura4⁺* transgene at the centromere. The two restriction fragments were analysed using a probe to origin 2055 (ref. 20) and a probe that recognizes the *ura4-DS/E* minigene deleted region. See oligonucleotide list for details (Supplementary Table 3).

Hydroxyurea synchronization and chromatin immunoprecipitation. Synchronization of log-phase *rad22-yfp* and *rad22-yfp/Adcr1* *S. pombe* strains

was performed as previously described². In brief, cells were treated with 15 mM hydroxyurea (Sigma) for 4.5 h to synchronize in S phase. Cells were released by washing twice in hydroxyurea-free media and grown for 5 h taking 50 ml samples every 30 min. Samples were fixed with 3% paraformaldehyde for ChIP as previously described⁷. Chromatin was purified, sonicated and incubated with either anti-RNA Polymerase II (clone 8WG16, Upstate) or anti-green fluorescent protein (Roche) antibodies. Synchronization efficiency was estimated using the septation index. For chromatin immunoprecipitation quantification, immunoprecipitated DNA was amplified by quantitative PCR (qPCR) in duplicate. The following formula was used: enrichment = $2^{-((C_{\text{test}} - C_{\text{act1}})_{\text{IP}} - (C_{\text{test}} - C_{\text{act1}})_{\text{wce}})}$ where C_{test} and C_{act1} is the amplification cycle in which the signal surpassed the established threshold for the test and reference amplicons respectively, in the immunoprecipitated (IP) and input DNA (wce) samples. The test regions were amplified with the following primers; p30F_qPCR and p30R_qPCR (*dh* region), p33F_qPCR and p33R_qPCR (*dg* region), and p20F and p20R (2055 origin of replication), and for the reference, act1F_qPCR and act1R_qPCR (*act1*). See oligo list for details (Supplementary Table 3).

Chromatin immunoprecipitation sequencing and data analysis. Samples from log-phase DG21 and DG690 *S. pombe* strains were fixed (dividing cells) or first treated with 15 mM hydroxyurea (Sigma) for 4.5 h, washed once in hydroxyurea-free media and fixed (hydroxyurea-arrested cells)². Chromatin was purified, sonicated and incubated with following anti-RNA polymerase II antibodies from Abcam: clone pospho-S2 (ab5095), and clone pospho-S5 (ab5131). For ChIP-seq, immunoprecipitated DNA was polished, ligated and amplified with the Illumina ChIP-seq sample preparation kit according to the manufacturer's instructions, and sequenced in an Illumina Hi-Seq 2000 Analyzer by 50 bp paired end sequencing as previously described³³. Analysis was carried out by a custom pipeline using Bowtie to align reads to the *S. pombe* genome and MACS 1.4 for significance enrichment calls³⁴. Wiggle fragment pileup files were generated for each comparison to visualize the difference in Pol II enrichment between control (DG21) and treatment (DG690) immunoprecipitations at 10-bp intervals genome wide. The difference at each interval was calculated by first normalizing to the total number of reads between the two by scaling up, and then subtracting the normalized control value from the treatment. Detailed sequence statistics are provided in Supplementary Table 4.

Microscopy. Samples were analysed using differential interference contrast (DIC) microscopy and fluorescence microscopy (Axioimager upright fluorescent microscope).

Survival assays. Mid log-phase cultures were resuspended to 2×10^6 cells per ml and serially diluted tenfold. Dilutions were spotted on YES agar plates or YES agar containing the indicated amounts of hydroxyurea (Sigma), camptothecin (Sigma) or methyl methanesulphonate (MMS) (Sigma). Recovery was for 3 days at 30 °C.

Co-immunoprecipitation. Cells were lysed by glass beads in HB buffer³¹. Lysates were pre-cleared with protein A agarose beads, followed by 2-h incubation with antibodies against the haemagglutinin tag (Sigma, E6779) at 4 °C. After washing, eluted proteins and input extracts were analysed by western blotting using antibodies against TAP (Sigma, P 2026).

- Moreno, S., Klar, A. & Nurse, P. Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194**, 795–823 (1991).
- Sallés, F. J., Richards, W. G. & Strickland, S. Assaying the polyadenylation state of mRNAs. *Methods* **17**, 38–45 (1999).
- Zaratigui, M. et al. CENP-B preserves genome integrity at replication forks paused by retrotransposon LTR. *Nature* **469**, 112–115 (2011).
- Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).